Beyond Riemannian geometry: The affine connection setting for transformation groups

Xavier Pennec, Marco Lorenzi

Beyond Riemannian Geometry

The affine connection setting for transformation groups

Xavier Pennec[a,*] and Marco Lorenzi[*]
* Université Côte d'Azur and Inria, Epione team, 2004 Route des Lucioles, BP93, F-06902 Sophia-Antipolis Cedex, France
[a] Corresponding: xavier.pennec@inria.fr

Abstract

In computational anatomy, one often lifts the statistics from the object space (images, surfaces, etc) to the group of deformation acting on their embedding space. Statistics on transformation groups have been considered in previous chapters, by providing the Lie group with a left or right invariant metric, which may (or may not) be consistent with the group action on our original objects. In this chapter we take the point of view of statistics on abstract transformations, independently of their action. In this case, it is reasonable to ask that our statistical methods respect the geometric structure of the transformation group. For instance, we would like to have a mean which is stable by the group operations (left and right composition, inversion). Such a property is ensured for Fréchet means in Lie groups endowed with a bi-invariant Riemannian metric, like compact Lie groups (e.g. rotations). Unfortunately bi-invariant Riemannian metrics do not exist for most non compact and non-commutative Lie groups, including rigid-body transformations in any dimension greater than one. Thus, there is a need for the development of a more general non-Riemannian statistical framework for general Lie groups.

In this chapter, we partially extend the theory of geometric statistics developed in the previous chapters to affine connection spaces. More particularly, we consider connected Lie groups endowed with the canonical Cartan-Schouten connection (a generally non-metric connection). We show that this connection provides group geodesics which are completely consistent with the composition and inversion. With such a non-metric structure, the mean cannot be defined by minimizing the variance as in Riemannian Manifolds. However, the characterization of the mean as an exponential barycenter gives us an implicit definition of the mean using a general barycentric equation. Thanks to the properties of the canonical Cartan connection, this mean is naturally bi-invariant. In finite dimension, this provides strong theoretical bases for the use of one-parameter subgroups. The generalization to infinite dimensions is at the basis of the SVF-framework. From the practical point of view, we show that it leads to efficient and plausible models of atrophy of the brain in Alzheimer's disease.

Author preprint. Chapter 5 of *Riemannian Geometric Statistics in Medical Image Analysis, Pages 169-229.*
X. Pennec, S. Sommer and T. Fletcher editors. Elsevier, 2020. With authorization.
https://doi.org/10.1016/B978-0-12-814725-2.00012-1

5.1. Introduction

In computational anatomy, one needs to perform statistics on shapes and transformations, and to transport these statistics from one point to another (e.g. from a subject to a template or to another subject). In this chapter we consider the points of view of abstract transformations, independently of their action on objects. To perform statistics on such transformation groups, the methodology developed in the previous chapters consists in endowing the Lie group with a left (or right) invariant metric which turns the transformation group into a Riemannian manifold. Then, one can use the tools developed in chapter 2 such as the Fréchet mean, tangent PCA or PGA.

On a Lie group, this Riemannian approach is consistent with the group operations if a bi-invariant metric exists, which is for example the case for compact groups such as rotations. In this case, the bi-invariant Fréchet mean has many desirable invariance properties: it is invariant with respect to left- and right-multiplication, as well as to inversion. However, a left-invariant metric on a Lie groups is generally not right-invariant (and conversely). Since the inverse operation exchanges a left-invariant metric for a right-invariant one, such metrics are generally not inverse consistent either. In this case, the Fréchet mean based on a left-invariant distance is not consistent with right translations nor with inversions. A simple example of this behavior is given in Section 5.5.2.4 with 2D rigid-body transformations.

In parallel to Riemannian methods based on left or right-invariant metrics, numerous methods in Lie groups are based on the group properties, and in particular on one-parameter subgroups, realized by the matrix exponential in matrix Lie groups. There exist particularly efficient algorithms to compute the matrix exponential like the scaling and squaring procedure [Hig05] or for integrating differential equations on Lie groups in geometric numerical integration theory [HLW02, IMKNZ00]. In medical image registration, parametrizing diffeomorphism with the flow of Stationary Velocity Fields (SVF) was proposed in [ACPA06] and very quickly adopted by many other authors. The group structure was also used to obtain efficient polyaffine transformations in [ACAP09]. Last but not least, [APA06, PA12] showed that the bi-invariant mean could be defined on Lie groups when the square-root (thus the log) of the transformations exists.

The goal of this chapter is to explain the mathematical roots of these algorithms, which are actually based on an affine connection structure instead of a Riemannian one. The connection defines the parallel transport, and thus a notion of geodesics that extends straight lines: the tangent vector to an auto-parallel curve stays parallel to itself all along the curve. In finite dimension, these geodesics define exponential and log maps locally, so we may generalize some of the statistical tools developed in Chapter 2 to affine connection spaces. As there is no distance, the Fréchet / Karcher means have to be replaced by the weaker notion of exponential barycenters (which are the critical points of the variance in Riemannian manifolds). Higher order moments and the Mahalanobis distance can be defined (locally) without trouble. However, tangent PCA is not easy to generalize as there is no reference metric to diagonalize the covariance matrix.

In the case of Lie groups, we describe in Section 5.3 the natural family of connections proposed by Cartan and Schouten in 1926 which are left and right-invariant, and for which one-parameter subgroups (the flow of left-invariant vector fields) are the geodesics going through the identity. Among these, there is a unique symmetric (torsion free) connection, called the canonical symmetric Cartan-Schouten (CCS) connection. We show in Section 5.4 that when there exists a bi-invariant metric on the Lie group (i.e. when the group is the direct product of Abelien and compact groups), the CCS connection is the Levi-Civita connection of that metric. However, the CCS connection still exists even when there is no bi-invariant metric. This is also the unique affine connection induced by the canonical symmetric space structure of the Lie groups with the symmetry $s_g(h) = gh^{-1}g$.

Based on the bi-invariance properties of the CCS connection, we turn in Section 5.5 to bi-invariant group means defined as exponential barycenters. These means exist locally and are unique for sufficiently concentrated data. In a number of Lie groups which do not support any bi-invariant metric, for instance nilpotent or some specific solvable groups, we show that there is even global uniqueness. Thus, the group mean appears to be a very general and natural notion on Lie groups. However, the absence of a metric significantly complexifies the analysis to identify whenever data are sufficiently concentrated or not, contrarily to Riemannian manifolds where we now have fairly tight conditions on the support of the distribution to ensure the existence and uniqueness of the Riemannian Fréchet mean. The particular case of rigid-body transformations shows that fairly similar conditions could be derived for the bi-invariant group mean. The question of how this result extends to other Lie groups remains open.

Section 5.6 investigates the application of the CCS affine connection setting to Lie groups of diffeomorphisms, which justifies the use of deformations parametrized by the flow of stationary velocity fields (SVF) in medical image registration. Although a number of theoretical difficulties remains when moving to infinite dimensions, very efficient algorithms are available, like the scaling and squaring to compute the group exponential and its (log)-Jacobian, the composition using the Baker-Campbell-Hausdorff formula, etc. This allows the straightforward extension of the classical "demons" image registration algorithm to encode diffeomorphic transformations parametrized by SVFs. A special feature of the log-demons registration framework is to enforce almost seamlessly the inverse consistency of the registration.

The log-demons algorithm can be used to accurately evaluate the anatomical changes over time for different subjects from time-sequences of medical images. However, the resulting deformation trajectories are expressed in each subject's own geometry. Thus, it is necessary to transport the intra-subject deformation information to a common reference frame in order to evaluate the differences between clinical groups in this type of longitudinal study. Instead of transporting a scalar summary of the deformation trajectory (usually the (log)-Jacobian determinant), we can transport the parameters of the full deformation thanks to the parallel transport. We detail in Section 5.7 two discrete parallel transport methods based on the computation of the geodesics: the Schild's ladder, and a more symmetric variant called the pole ladder, specialized for the parallel transport along geodesics. Most parallel transport methods are shown to be first order approximations. It is noticeable that the pole ladder is actually a third order scheme which becomes exact in symmetric spaces. We illustrate the computational advantages and demonstrate the numerical accuracy of this method by providing an application to the modeling of the longitudinal atrophy progression in Alzheimer's disease (AD). The quality of the resulting average models of normal and disease deformation trajectories suggests that an important gain in sensitivity could be expected in group-wise comparisons. For example, a statistically significant association between the brain shape evolution and the $A\beta_{1-42}$ biomarker among normal subjects of the ADNI study was identified in [LFAP11]. Moreover, this framework is among the leading methods benchmarked in [CFI+15] for the quantification of longitudinal hippocampal and ventricular atrophy, showing favorable results compared to state-of-the-art approaches based on image segmentation.

5.2. Affine Connections Spaces

Geodesics, exponential and log maps are among the most fundamental tools to work on differential manifolds, as we have seen in the previous chapters. However, when our manifold is given an additional structure, which is the case for instance for Lie groups, a compatible Riemannian metric does not always exist. We investigate in this section how we can define the notion of geodesics in non-Riemannian spaces. The main idea is to rely on the notion of curve with vanishing acceleration, or equivalently curves whose tangent vectors remains parallel to themselves (auto-parallel curves). In order to compare vectors living in different tangent spaces (even at points which are infinitesimally close), we need to provide a notion of parallel transport from one tangent space to the other. Likewise, computing accelerations requires a notion of infinitesimal parallel transport that is called a connection.

We consider a smooth differential manifold $\mathcal{M}$. We denote by $T_p\mathcal{M}$ the tangent space at $p \in \mathcal{M}$ and by $T\mathcal{M}$ the tangent bundle. A section of that bundle $X : \mathcal{M} \mapsto T\mathcal{M}$ is a smooth vector field whose value at p is denoted $X|_p$. The set of vector fields, denoted $\Gamma(\mathcal{M})$, can be identified with derivations. Recall that a derivation δ is a linear map from $\Gamma(\mathcal{M})$ to itself that satisfies the Leibniz's law: $\delta(fX) = (df)X + f(\delta X)$ for any $f \in C^\infty(\mathcal{M})$ and $X \in \Gamma(\mathcal{M})$. In a local coordinate system, a derivation writes as $X\phi|_p = \partial_X\phi|_p = \frac{d}{dt}\left(\phi(p + tX|_p)\right)$. In this chapter, we prefer the notation $\partial_X\phi$ to the notation $X\phi$. Composing two derivations is in general not a derivation because a term with second order derivative appears when we write it in a local coordinate system. However, this second order term can be canceled by subtracting the symmetric composition, so that the bracket $[X, Y](\phi) = \partial_X\partial_Y\phi - \partial_Y\partial_X\phi$ of two vector fields is itself a vector field. Endowed with this bracket, the set of vector fields $\Gamma(\mathcal{M})$ is the algebra of derivations of smooth functions $\phi \in C^\infty(\mathcal{M})$.

5.2.1. Affine Connection as an Infinitesimal Parallel Transport

In order to compare vectors in the tangent space at one point of the manifold with vectors that are tangent at a different point, we need to define a mapping between these two tangent spaces: this is the notion of parallel transport. As there is generally no way to define globally a linear operator $\Pi_p^q : T_p\mathcal{M} \to T_q\mathcal{M}$ which is consistent with the composition (i.e. $\Pi_p^q \circ \Pi_r^p = \Pi_r^q$), the path connecting the two points p and q has to be specified.

Definition 1 (Parallel transport along a curve)**.** Let γ be a curve in $\mathcal{M}$ joining $\gamma(s)$ to $\gamma(t)$. A parallel transport assigns to each curve a collection of mappings $\Pi(\gamma)_s^t : T_{\gamma(s)}\mathcal{M} \to T_{\gamma(t)}\mathcal{M}$ such that:

- $\Pi(\gamma)_s^s =$ Id, the identity transformation of $T_{\gamma(s)}\mathcal{M}$.
- $\Pi(\gamma)_u^t \circ \Pi(\gamma)_s^u = \Pi(\gamma)_s^t$ (consistency along the curve).
- The dependence of Π on γ, s, and t is smooth.

The notion of (affine) connection, also called covariant derivative, is the infinitesimal version of the parallel transport for the tangent bundle. Let $X|_p = \dot\gamma(0)$ be the tangent vector at the initial point p of the curve γ, and Y be a vector field. Then $\nabla_X Y = \frac{d}{dt}\Pi(\gamma)_t^0 Y|_{\gamma(t)}\Big|_{t=0}$ defines a bilinear map which is independent of the curve γ and only depends on the vector fields X and Y.

Definition 2 (Affine connection)**.** An affine connection (or covariant derivative) on the manifold $\mathcal{M}$ is a bilinear map $(X, Y) \in \Gamma(\mathcal{M}) \times \Gamma(\mathcal{M}) \mapsto \nabla_X Y \in \Gamma(\mathcal{M})$ such that for all smooth functions $\phi \in C^\infty(\mathcal{M})$:

- $\nabla_{\phi X} Y = \phi\nabla_X Y$, that is, ∇ is smooth and linear in the first variable;
- $\nabla_X(\phi Y) = \partial_X\phi Y + \phi\nabla_X Y$, that is, ∇ satisfies Leibniz rule in the second variable.

In a local chart, the vector fields ∂_i constitute a basis of $\Gamma(\mathcal{M})$: a vector field $X = x^i\partial_i$ has coordinates $x^i \in C^\infty(\mathcal{M})$ (using Einstein summation convention). Similarly let $Y = y^i\partial_i$. Using the two above rules, we can write the connection:

$$\nabla_X Y = x^i \nabla_{\partial_i}(y^j \partial_j) = x^i y^j \nabla_{\partial_i}\partial_j + x^i \partial_i y^j \partial_j = x^i y^j \nabla_{\partial_i}\partial_j + \partial_X Y.$$

This means that the connection is completely determined by its coordinates on the basis vector fields $\nabla_{\partial_i}\partial_j = \Gamma_{ij}^k \partial_k$. The n^3 coordinates Γ_{ij}^k of the connection are called the Christoffel symbols. They encode how the transport from one tangent space to neighboring ones modifies the standard derivative of a vector field in a chart: $\nabla_X Y = \partial_X Y + x^i y^j \Gamma_{ij}^k \partial_k$.

Definition 3 (Covariant derivative along a curve)**.** Let $\gamma(t)$ be a curve on $\mathcal{M}$ and $Y = Y(\gamma(t))$ be a vector field along the curve. The covariant derivative of Y along $\gamma(t)$ is

$$\frac{\nabla Y}{dt} = \nabla_{\dot\gamma} Y = \partial_{\dot\gamma} Y + Y^i \dot\gamma^j \Gamma_{ij}^k(\gamma)\partial_k.$$

A vector field is covariantly constant along the curve γ if $\nabla Y/dt = \nabla_{\dot\gamma} Y = 0$. This defines the parallel transport according to the connection: specifying the value of the field at one point $Y(\gamma(0)) = u$ determines the value $Y(\gamma(t)) = \Pi(\gamma)_0^t u$ of a covariantly constant field at all the other point of the curve.

The connection we defined so far is differentiating vector fields. It can be uniquely extended to covectors and more general tensor fields by requiring the resulting operation to be compatible with tensor contraction and the product rule $\nabla_X(Y \otimes Z) = (\nabla_X Y) \otimes Z + Y \otimes (\nabla_X Z)$.

5.2.2. Geodesics

Looking for curves whose tangent vectors are covariantly constant provides a definition of geodesics in affine connection spaces that generalizes straight lines: these are the curves that remain parallel to themselves (auto-parallel curves).

Definition 4 (Affine geodesics)**.** $\gamma(t)$ is a geodesic if its tangent vector $\dot\gamma(t)$ remains parallel to itself, i.e. if the covariant derivative $\nabla_{\dot\gamma}\dot\gamma = 0$ of γ is zero. In a local coordinate system, the equation of the geodesics is thus: $\ddot\gamma^k + \Gamma_{ij}^k \dot\gamma^i \dot\gamma^j = 0$.

We retrieve here the standard equation of the geodesics in Riemannian geometry without having to rely on any particular metric. However, what is remarkable is that we still conserve many properties of the Riemannian exponential map in affine connection spaces: as geodesics are locally defined by a second order ordinary differential equation, the geodesic $\gamma_{(p,v)}(t)$ starting at any point p with any tangent vector v is defined for a sufficiently small time, which means that we can define the affine exponential map $\exp_p(v) = \gamma_{(p,v)}(1)$ for a sufficiently small neighborhood. Moreover, the *strong*

Whitehead theorem still holds.

Theorem 5.1 (Strong Form of Whitehead Theorem)**.** *Each point of an affine connection space has a* normal convex neighborhood *(NCN) in the sense that for any couple of points (p, q) in this neighborhood, there exists a unique geodesic $\gamma(t)$ joining them that is entirely contained in this neighborhood. Moreover, the geodesic $\gamma(t)$ depends smoothly on the points p and q.*

The proof of this theorem essentially involves the non-singularity of the differential of the map $\Phi(p, v) = (p, \exp_p(v))$ and the inverse function theorem, with the use of an auxiliary Euclidean metric on the tangent spaces around the point of interest. We refer to [Pos01, Proposition 1.3, p.13] for the detailed proof.

As geodesics control many properties of the space, it is interesting to know which affine connections lead to the same geodesics. Intuitively, a geodesic for a connection ∇ will remain a geodesic for another connection $\bar{\nabla}$ if the parallel transport of the tangent vector *in the direction of this tangent vector* is the same, i.e. if $\nabla_X X = \bar{\nabla}_X X$ for any vector field X. However, the parallel transport of other vectors of a frame can change: this can be measured by torsion. The curvature is a second order measure of how the parallel transport differ along different paths.

Definition 5 (Torsion and curvature tensors of an affine connection)**.** The torsion of an affine connection is:

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] = -T(Y, X).$$

This tensor measures how the skew-symmetric part differ from the Lie derivative $\mathcal{L}_X Y = [X, Y]$. The connection is called torsion free (or symmetric) if the torsion vanishes. The curvature tensor is defined by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z.$$

It measures how the infinitesimal parallel transport differ along the sides of a geodesic parallelogram.

One can show that two connections have the same geodesics if they have the same symmetric part $(\nabla_X Y + \nabla_Y X)/2$. i.e. if they only differ by torsion. Thus, at least for the geodesics, we can restrict our attention to the torsion free connections. Conversely, any procedure involving only geodesics (such as the Schild's or the pole ladders developed later in this chapter) is insensitive to the torsion of a connection.

5.2.3. Levi Civita Connection of a Riemannian metric

Now that we have developed the theory of affine connection spaces, it is time to see how it relates to Riemannian geometry. Let $g_{ij} = \left\langle \partial_i, \partial_j \right\rangle$ be a smooth positive definite bilinear symmetric form on $T\mathcal{M}$. With this metric, we can define geodesics which are (locally) length minimizing. Is there a specific connection for which the two notions of geodesics are the same? The answer is yes, the two constructions do agree thanks to the fundamental theorem of Riemannian geometry.

Theorem 5.2 (Levi-Civita connection of a Riemannian manifold)**.** *On any Riemannian manifold, there exists a unique torsion free connection which is compatible with the metric as a derivation, called the* Levi-Civita connection*. It satisfies:*

- $T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] = 0$ *(symmetry).*
- $\partial_X \langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$ *(compatibility with the metric).*

The proof is constructive. First, we expand $\partial_X \langle Y, Z\rangle + \partial_Y \langle X, Z\rangle - \partial_Z \langle X, Y\rangle$ using the metric compatibility condition. Then, we use the zero torsion formula to replace $\nabla_Y X$ by $\nabla_X Y + [Y, X]$ and similarly for $\nabla_Z Y$ and $\nabla_Z X$. We obtain Khozul formula, which uniquely defines the scalar product of the connection with any vector field (thus the connection):

$$2\langle \nabla_X Y, Z\rangle = \partial_X \langle Y, Z\rangle + \partial_Y \langle X, Z\rangle - \partial_Z \langle X, Y\rangle + \langle [X, Y], Z\rangle - \langle [X, Z], Y\rangle - \langle [Y, Z], X\rangle$$

Written in a local coordinate system, we have $\left\langle \nabla_{\partial_i} \partial_j, \partial_k \right\rangle = g_{mk} \Gamma^m_{ij}$. Let $[g^{ij}] = [g_{ij}]^{-1}$ be the inverse of the metric matrix. Then the extension of the right part of Khozul formula shows that the Christoffel symbols of the Levi-Civita connection are

$$\Gamma^i_{jk} = \frac{1}{2} g^{im} \left(\partial_k g_{mj} + \partial_j g_{mk} - \partial_m g_{jk} \right). \tag{5.1}$$

5.2.4. Statistics on Affine Connection Spaces

In a connected manifold of dimension d which is orientable and countable at infinity (we don't even need the affine structure), the bundle of smooth d-forms $\Omega^d(\mathcal{M}) = \Lambda^d T^*\mathcal{M}$ is trivial so that there exist sections that never vanish (volume forms). Focusing on non-negative forms that sum-up to 1, we get the space of probabilities $\mathrm{Prob}(M) = \left\{ \mu \in \Omega^d(\mathcal{M}) \mid \int_M \mu = 1,\ \mu \geq 0 \right\}$. With this definition, the space of probabilities is not a manifold but a stratified space because we hit a boundary at each point where $\mu(dx) = 0$. Imposing $\mu > 0$ gives the smooth manifold of never vanishing probabilities that can be endowed with the canonical Fisher-Rao metric, which is the unique metrique invariant under the action of diffeomorphisms (reparametrizations of $\mathcal{M}$) for compact manifolds [BBM14]. However, we do not need this metric structure here, so it is more interesting to consider the space $\mathrm{Prob}(M)$ comprising both smooth probabilities and sample distributions like $\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}$.

As for Riemannian manifolds, we cannot define the mean value in an affine connection space with an expectation, unless the space is linear. Since we don't have a distance, we cannot use the minimization of the expected square distance either (the Fréchet mean). However, we may rely on one of their properties: the implicit localization of the mean as a barycenter in exponential coordinates. The weaker affine structure makes it challenging to determine the existence and uniqueness conditions of this more general definition. However, this can be worked out locally. We can also define higher order moments and some simple statistical tools like the Mahalanobis distance. Other notions like principal component analysis cannot be consistently defined without an additional structure. In the generalization of statistics from Riemannian manifolds to affine connection spaces that we investigate in this section, an important issue is to distinguish the affine notions (relying on geodesics) from the Riemannian ones (relying on the distance).

Mean Values with Exponential Barycenters

Let us start by recalling the classical definition of a mean in the Euclidean space $\mathbb{R}^d$: the mean (or **barycenter**) of a probability μ (or a set of points $\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{p_i}$) is the unique point p that verifies the barycentric equation $\int_{\mathbb{R}^d}(p-q)\,\mu(dq) = 0$. At the mean, the sum of the weighted displacements to each of the sample points is null, i.e. the mean is at the center of the data. Notice that this definition is affine because it does not require the dot product.

In a manifold with Riemannian metric $\|.\|_p$, the Fréchet mean of a probability μ is the set of minima of the variance $\sigma^2(p) = \int_{\mathcal{M}} \mathrm{dist}(p,q)^2\,\mu(dq) = \int_{\mathcal{M}} \|\mathrm{Log}_p(q)\|_p^2\,\mu(dq)$. When the cut locus has null measure $\mu(Cut(p)) = 0$, the variance is differentiable and its gradient is $\nabla\sigma^2(p) = -2\int_{\mathcal{M}} \mathrm{Log}_p(q)\,\mu(dq) = -2\mathfrak{M}_1(\mu)$. The local minima being critical points, they satisfy the barycentric equation $\int_{\mathcal{M}} \mathrm{Log}_p(q)\,\mu(dq) = 0$ when they belong to the punctured manifold $\mathcal{M}_\mu^* = \{p \in \mathcal{M} \mid \mu(Cut(p)) = 0\}$. This critical condition equation can also be taken as the definition of the mean, which leads to the notion of exponential barycenter proposed in [BK81, EM91, Arn94, Arn95, CK99].

In an affine connection space, the notion of exponential barycenter still makes sense as soon as we can compute the affine logarithm. The Whitehead Theorem (5.1) tells us that there exists a normal convex neighborhood U at each point of the manifold. For sufficiently concentrated probabilities $\mu \in \mathrm{Prob}(U)$ with support in such a neighborhood U, computing the first moment field $\mathfrak{M}_1(\mu) = \int_{\mathcal{M}} \log_p(q)\,\mu(dq)$ makes sense for $p \in U$.

Definition 6 (Exponential barycenter in an affine connection space)**.** For sufficiently concentrated probabilities $\mu \in \mathrm{Prob}(U)$ whose support is included in a normal convex neighborhood U, exponential barycenters (of that neighborhood) are the points implicitly defined by

$$\mathfrak{M}_1(\mu)|_p = \int_{\mathcal{M}} \log_p(q)\, \mu(dq) = 0. \tag{5.2}$$

This definition is close to the Riemannian center of mass (or more specifically the Riemannian average) of [Gro04] but uses the logarithm of the affine connection instead of the Riemannian logarithm. Notice however that there is an implicit dependency of the convex neighborhood considered in that definition. In the absence of an auxiliary metric, there is no evident way to specify a natural convex neighborhood which might be maximal.

Theorem 5.3 (Existence of exponential barycenters)**.** *Distributions with compact support in a normal convex neighborhood U of an affine manifold (M, ∇) have at least one exponential barycenter. Moreover, exponential barycenters are stable by affine diffeomorphisms (connection preserving maps, which thus preserve the geodesics and the normal convex neighborhoods).*

Proof. The skecth of the proof of [PA12] is the following. First the normal neighborhood U being star-shaped, it is homeomorphic to a sphere. Second, because U is convex, the convex combination of vectors of $\log_p(q) \in \log_p(U)$ remains in $\log_p(U)$ so that geodesic shooting along $\mathfrak{M}_1(\mu)$ remains in the normal convex neighborhood: $\exp_p(\mathfrak{M}_1(\mu)) \in U$. By Brouwer's fixed-point theorem, there is thus a fixed point within U verifying $\exp_p(\mathfrak{M}_1(\mu)) = p$, or equivalently $\mathfrak{M}_1(\mu) = \int_{\mathcal{M}} \log_p(q)\, \mu(dq) = 0$. A slightly different proof of existence using the index of the vector field $\mathfrak{M}_1(\mu)$ is due to [BK81]. See also this reference for the stability under affine maps. □

Remark 1. For distributions whose support is too large, exponential barycenters may not exist. One reason is that the affine logarithm might fail to exist, even for geodesically complete affine connection manifolds. The classical example is the group of unimodular real matrices of dimension two $\mathrm{SL}(2, \mathbb{R}) = \{A \in \mathrm{GL}(2, \mathbb{R}) \mid \det(A) = 1\}$ endowed with the canonical symmetric connection (see next section) for which one can show that a matrix $A \in \mathrm{SL}(2, \mathbb{R})$ has the form $\exp(X)$ iff $A = -\mathrm{Id}_2$ or $\mathrm{Tr}(A) > -2$. Thus, matrices with $\mathrm{Tr}(A) < -2$ can't be reached by any geodesic starting from identity. As a consequence, a distributions on $\mathrm{SL}(2, \mathbb{R})$ with some mass in this domain and at the identity do not fit in a normal convex neighborhood. Such a distribution has no exponential barycenter. We should relate this failure of existence of the mean to a classical phenomenon observed in Euclidean spaces with heavy tailed distributions, where the first moment is unbounded. It is currently not known if bounding the first moments (with respect to an auxiliary norm) is a sufficient condition for the existence of exponential barycenters in affine connection spaces.

Uniqueness conditions of the exponential barycenter are harder to determine. There exist convex manifolds (affine connection spaces restrained to a convex neighborhood) which do not have a unique exponential barycenter. This is the case of the propeller manifold of [Ken92] where a sample probability of three points with four exponential barycenters is explicitly constructed. Under additional assumptions controlling the covariant derivative of the logarithm inherited from Riemannian manifolds, [BK81] could show the uniqueness of the exponential barycenter. Generalizing these curvature-based conditions to general affine connection spaces remains to be done.

A different approach relying on a stronger notion of convexity was proposed by Arnaudon and Li in [AL05].

Definition 7 (p-convexity)**.** Let $(\mathcal{M}, \nabla)$ be an affine manifold. A separating function on $\mathcal{M}$ is a convex function $\phi : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^+$ vanishing exactly on the diagonal of the product manifold (considered as an affine manifold with the direct product connection). Here, convex means that the restriction of $\phi(\gamma(t))$ to any geodesic $\gamma(t)$ is a convex function from $\mathbb{R}$ to $\mathbb{R}^+$. A manifold which carries a smooth separating function ϕ such that

$$c\, d(p, q)^p \leq \phi(p, q) \leq C\, d(p, q)^p,$$

for some constants $0 < c < C$, some positive integer p and some Riemannian distance function d is called a manifold with p-convex geometry.

A separating function generalizes the separating property of a distance ($\phi(p, q) = 0 \Leftrightarrow p = q$) and the ever increasing value when q gets away from p along geodesics but does not respect the triangular inequality. Whitehead Theorem tells us in essence that any point in an affine connection space has a convex neighborhood with 2-convex geometry. In Riemannian manifolds, sufficiently small geodesic balls have 2-convex geometry. p-convexity (for $p \geq 2$) is sufficient to ensure uniqueness but there are examples of manifolds where one can prove uniqueness although they do not have p-convex geometry for any p. This is the case of open hemispheres endowed with the classical Euclidean embedding metric even if each geodesic ball strictly smaller than an open hemisphere has p-convex geometry for some p depending on the radius [Ken91]. This motivates the following extension.

Theorem 5.4 (Uniqueness of the exponential barycenter in CSLCG Manifolds [AL05])**.** *A convex affine manifold* (M, ∇) *is said to be CSLCG (convex, with semi-local convex* *geometry) if there exists an increasing sequence $(U_n)_{n\geq 1}$ of relatively compact open convex subsets of $\mathcal{M}$ such that $\mathcal{M} = \cup_{n\geq 1} U_n$ and for every U_n ($n \geq 1$), U_n has p-convex geometry for some $p \in 2\mathbb{N}$ depending on n. On a CSLCG manifold, every probability measure with compact support has a unique exponential barycenter.*

Covariance Matrix and Higher Order Moments

The mean is an important statistic which indicates the location of the distribution in the group, but higher order moments are also needed to characterize the dispersion of the population around this central value. Exponential barycenters are defined above for sufficiently concentrated probabilities $\mu \in \text{Prob}(U)$ with support in a normal convex neighborhood U, so that computing the first moment field $\mathfrak{M}_1(\mu) = \int_{\mathcal{M}} \log_p(q)\, \mu(dq)$ make sense for $p \in U$. In this neighborhood, computing higher order moments also make sense.

Definition 8 (Moments of a probability distribution in an affine connection space)**.** For sufficiently concentrated probabilities $\mu \in \text{Prob}(U)$ whose support is included in a normal convex neighborhood U, the k-order moment is the k-contravariant tensor:

$$\mathfrak{M}_k(\mu)|_p = \int_{\mathcal{M}} \underbrace{\log_p(q) \otimes \log_p(q) \ldots \otimes \log_p(q)}_{k \text{ times}} \mu(dq). \tag{5.3}$$

In particular, the covariance field is the 2-contravariant tensor with the following coordinates in any basis of the tangent space $T_p\mathcal{M}$:

$$\Sigma|_p^{ij} = \int_{\mathcal{M}} [\log_p(q)]^i [\log_p(q)]^j \mu(dq),$$

and its value $\Sigma = \Sigma|_{\bar{p}}$ at the exponential barycenter $\bar{p}$ solution of $\mathfrak{M}_1(\mu)|_{\bar{p}} = 0$ (assuming it is unique) is called the covariance of μ.

One should be careful that the covariance defined here is a rank $(2, 0)$ (2-contravariant) tensor which is not equivalent to the usual covariance matrix seen as a bilinear form (a 2-covariant or rank $(0, 2)$ tensor belonging to $T^*_{\bar{p}}\mathcal{M} \otimes T^*_{\bar{p}}\mathcal{M}$), unless we have an auxiliary metric to lower the indices. In particular, the usual interpretation of the coordinates of the covariance matrix using the scalar products of data with the basis vectors $\Sigma_{ij} = \int_{\mathcal{M}} < \log_p(q), e_i > < \log_p(q), e_j > \ \mu(dq)$ is valid only if we have a local metric $< ., . >$ which also defines the orthonormality of the basis vectors e_i. Likewise, diagonalizing Σ to extract the main modes of variability only makes sense with respect to a local metric: changing the metric of $T_{\bar{p}}\mathcal{M}$ will not only change the eigenvectors and the eigenvalues but also potentially the order of the eigenvalues. This means that Principle Component Analysis (PCA) cannot be generalized to the affine connection space setting without an additional structure. This is the reason why PCA is sometimes called Proper Orthogonal Decomposition (POD) in certain domains.

Mahalanobis Distance

Despite the absence of a canonical reference metric, some interesting tools can be defined from the 2-contravariant tensor in an intrinsic way without having to rely on an auxiliary metric. One of them is the Mahalanobis distance of a point q to a given distribution (in the normal convex neighborhood specified above):

$$d^2_{\mu}(q) = d^2_{(\bar{p},\Sigma)}(q) = [\log_{\bar{p}}(q)]^i \Sigma^{-1}_{ij} [\log_{\bar{p}}(q)]^j. \tag{5.4}$$

In this formula, $\bar{p}$ is the exponential barycenter of μ and Σ^{-1}_{ij} are the coefficients of the inverse of the covariance $\Sigma|^{ij}_{\bar{p}}$ in a given basis. One verifies that this definition does not depend on the basis chosen for $T_{\bar{p}}\mathcal{M}$. Furthermore, the Mahalanobis distance is invariant under affine diffeomorphisms of the manifold. Indeed, if ϕ is an affine diffeomorphism preserving the connection, then $U' = \phi(U)$ is a normal convex neighborhood that contain the support of the push-forward probability distribution $\mu' = \phi_* \mu$ and the differential $d\phi$ acts as a linear map on the tangent spaces: $\log_{\bar{\phi}(p)}(\phi(q)) = d\phi \log_{\bar{p}}(q)$. This shows that the exponential barycenter (assumed to be unique) is equivariant: $\bar{p}' = \overline{\phi(p)} = \phi(\bar{p})$ and that the covariance is transformed according to $\Sigma' = d\phi\, \Sigma\, d\phi^{\top}$ in the transformed coordinate system $e'_i = d\phi\, e_i$. Because of the inversion of the matrix Σ' in the Mahalanobis distance, we get the invariance property: $d^2_{(\bar{p}',\Sigma')}(q') = d^2_{(\bar{p},\Sigma)}(q)$, or more evidently $d^2_{\phi_*\mu}(\phi(q)) = d^2_{\mu}(q)$.

Open Problems for Generalizing Other Statistical Tool

This simple extension of the Mahalanobis distance suggests that it might be possible to extend much more statistical definitions and tools on affine connection spaces in a consistent way. One of the key problem is that everything we did so far is limited to small enough normal convex neighborhoods. To tackle distributions of larger non-convex support, we probably need an additional structure to specify a unique choice of a maximal domain for the log function, as this is done in the Riemannian case by the shortest distance criterion. A second difficulty relies in the absence of a natural reference measure on an affine connection space. This means that probability density functions have to be defined with respect to a given volume form, and would be different with the choice of another one. Thus, the notions of uniform distribution and entropy of a distribution need to be related to the chosen reference measure. This may considerably complexify the maximum entropy principle used in [Pen06a] and Chapter 3 to define the Gaussian distributions. However, certain specific geometric structures like Lie groups may provide additional tools to solve some of these problems, as we will see in the next sections.

5.3. Canonical Connections on Lie Groups

Let us come back now to Lie groups. We first recall a series of properties of Lie groups in Section 5.3.1 before turning to the search for affine connections that are compatible with the Lie group operations in Section 5.3.2.

5.3.1. The Lie Group Setting

Recall from chapter 1 that a Lie group G is a smooth manifold provided with an identity element e, a smooth composition rule $(g, h) \in G \times G \mapsto gh \in G$ and a smooth inversion rule $\mathrm{Inv} : f \mapsto f^{-1}$ which are both compatible with the manifold structure. The composition operation defines two canonical automorphisms of G called the left and the right translations: $L_g : f \mapsto gf$ and $R_g : f \mapsto fg$. The differential dL_g of the left translations maps the tangent space T_hG to the tangent space $T_{gh}G$. In particular, dL_g maps any vector $x \in T_eG$ to the vector $dL_gx \in T_gG$, giving rise to the vector field $\tilde{x}|_g = dL_g\, x$. One verifies that this vector field is left-invariant: $\tilde{x} \circ L_g = dL_g\tilde{x}$. Conversely, every left-invariant vector field is determined by its value at identity. Moreover, the bracket of two left-invariant vector fields $\tilde{x} = dL\, x$ and $\tilde{y} = dL\, y$ is also left-invariant and determined by the vector $[x, y] = [\tilde{x}, \tilde{y}]|_e \in T_eG$. Thus, left-invariant vector fields constitute a sub-algebra of the algebra of vector fields $\Gamma(G)$ on the group. This is a fundamental algebra for Lie groups which is called the *Lie algebra* $\mathfrak{g}$. As we have seen, the Lie algebra can be identified with the tangent vector space at identity, endowed with the bracket defined above. Thus, the Lie algebra has the same dimension than the group. Because any basis of the tangent space at identity is smoothly transported by left translation into a basis of the tangent space at any point, one can decompose any smooth vector field on the basis of left-invariant ones with coefficients that are smooth functions on the manifold. This means that the algebra of vector fields $\Gamma(G)$ is a finite dimensional module of dimension $dim(G)$ over the ring $C^\infty(G)$ of smooth functions over the group.

By symmetry, we can also define the sub-algebra of right-invariant vector fields $\bar{X}|_g = dR_gX$ and identify it with the tangent vector space at identity. However, one should be careful that the right-bracket defined on the tangent space at identity is the opposite of the left-bracket. This can be explained by the fact that left and right Lie algebras are related by the inversion map $\mathrm{Inv} : f \mapsto f^{-1}$ of the group, which exchanges left and right compositions, and whose differential at identity is $-\mathrm{Id}_{T_eG}$. As the algebra of right-invariant vector fields is usually used for diffeomorphisms instead of the traditional left-invariant ones for finite dimensional Lie groups, this explains why there is sometimes a minus sign in the bracket (see Section 5.6.2.4 and comments in [VPPA08, BHO07]).

Adjoint Group

A third important automorphism of G is the conjugation $C_g : f \mapsto gfg^{-1}$. Differentiating the conjugation with respect to f gives an automorphism of the Lie algebra called the *adjoint action*: an element g of G acts on an element x of $\mathfrak{g}$ by

$$\mathrm{Ad}(g)x = dL_g|_{g^{-1}} dR_{g^{-1}}|_e x = dR_{g^{-1}}|_g dL_g|_e x.$$

In the matrix case, we have the classical formula: $\mathrm{Ad}(B)M = BMB^{-1}$.

Thus, one can map each element of the group to a **linear operator** which acts on the Lie algebra. The mapping $\mathrm{Ad} : G \to \mathrm{GL}(\mathfrak{g})$ is moreover a Lie group homomorphism from G to $\mathrm{GL}(\mathfrak{g})$ since it is smooth and compatible with the group structure: $\mathrm{Ad}(e) = \mathrm{Id}, \forall g \in G,\ \mathrm{Ad}(g^{-1}) = \mathrm{Ad}(g)^{-1}$ and $\forall g, h \in G,\ \mathrm{Ad}(gh) = \mathrm{Ad}(g)\mathrm{Ad}(h)$. Thus, G can be 'represented' by the adjoint operators acting on $\mathfrak{g}$ in the sense of the representation theory (see [Lan04] for a complete treatment). The subgroup $\mathrm{Ad}(G)$ of the general linear group $\mathrm{GL}(\mathfrak{g})$ is called the **adjoint group**. The properties of this representation and the existence of bi-invariant metrics on the group G are intricated.

Taking the derivative of the adjoint map at the identity gives a representation of its Lie algebra: let $g(t)$ be a curve passing through the identity $g(0) = e$ with tangent vector $x \in \mathfrak{g}$. Then, for any $y \in \mathfrak{g}$, $d\mathrm{Ad}(g)y/dt = \mathrm{ad}x(y) = [x, y]$. Thus, $\mathrm{ad}x$ is a linear operator on the Lie algebra $\mathfrak{g}$.

Matrix Lie Group Exponential and Logarithm

In the matrix case, the exponential $\exp(M)$ of a square matrix M is given by $\exp(M) = \sum_{k=0}^{\infty} M^k/k!$. Conversely, let $A \in \mathrm{GL}(d)$ be an invertible sqaure matrix. If there exists a matrix $M \in M(d)$ such that $A = \exp(M)$, then M is said to be a logarithm of A. In general, the logarithm of a real invertible matrix may not exist, and it may not be unique if it exists. The lack of existence is an unavoidable phenomenon in certain connected Lie groups: one generally needs *two* exponentials to reach every element [Wö3]. When a real invertible matrix has no (complex) eigenvalue on the (closed) half line of negative real numbers, then it has a unique real logarithm whose (complex) eigenvalues have an imaginary part in $(-\pi, \pi)$ [KL89, Gal08]. In this case this particular logarithm is well-defined and called **the principal logarithm**. We will write $\log(M)$ for the principal logarithm of a matrix M whenever it is defined.

Thanks to their remarkable algebraic properties, and essentially their link with one-parameter subgroups, matrix exponential and logarithms can be quite efficiently numerically computed, for instance with the popular 'Scaling and Squaring Method' [Hig05] and 'Inverse Scaling and Squaring Method' [HCHKL01].

One Parameter Subgroups and Group Exponential

For general Lie groups, the flow $\gamma_x(t)$ of a left-invariant vector field $\tilde{x} = dL\, x$ starting from e with tangent vector x exists for all times. Its tangent vector is $\dot{\gamma}_x(t) = dL_{\gamma_x(t)} x$

by definition of the flow. Now fix $s \in \mathbb{R}$ and observe that the two curves $\gamma_x(s+t)$ and $\gamma_x(s)\gamma_x(t)$ are going through point $\gamma_x(s)$ with the same tangent vector. By the uniqueness of the flow, they are the same and γ_x is a one parameter subgroup, i.e. a group morphism from $(\mathbb{R}, 0, +)$ to $(G, e, .)$: $\gamma_x(s+t) = \gamma_x(s)\,\gamma_x(t) = \gamma_x(t+s) = \gamma_x(t)\,\gamma_x(s)$.

Definition 9 (Group exponential)**.** Let G be a Lie group and let x be an element of the Lie Algebra $\mathfrak{g}$. The **group exponential** of x, denoted $\exp(x)$, is given by the value at time 1 of the unique function $\gamma_x(t)$ defined by the ordinary differential equation (ODE) $\dot{\gamma}_x(t) = dL_{\gamma_x(t)}x$ with initial condition $\gamma_x(0) = e$.

One should be very careful that the group exponential is defined from the group properties only and does not require any Riemannian metric. It is thus generally different from the Riemannian exponential map associated to a Riemannian metric on the Lie group. However, both exponential maps share properties: in finite dimension, the group exponential is diffeomorphic locally around 0.

Theorem 5.5 (Group logarithm)**.** *In finite dimension, the group exponential is a diffeomorphism from an open neighborhood of* 0 *in* $\mathfrak{g}$ *to an open neighborhood of e in G, and its differential map at* 0 *is the identity. This implies that one can define without ambiguity an inverse map, called* **the group logarithm map** *in an open neighborhood of e: for every g in this open neighborhood, there exists a unique x in the open neighborhood of* 0 *in* $\mathfrak{g}$ *such that* $g = \exp(x)$.

Indeed, the exponential is a smooth mapping and its differential map is invertible at e. Thus, the Inverse Function Theorem guarantees that it is a diffeomorphism from some open neighborhood of 0 to an open neighborhood of $\exp(0) = e$ [Pos01, Proposition 1.3, p.13]. We write $x = \log(g)$ for this logarithm, which is the (abstract) equivalent of the (matrix) *principal* logarithm. Notice that we use lower case exp and log to clearly distinguish the group exponential and logarithm from their Riemannian counterparts Exp and Log. The absence of an inverse function theorem in infinite dimensional Fréchet manifolds prevents the straightforward extension of this property to general groups of diffeomorphisms [KW09].

Baker Campbell Hausdorff (BCH) Formula

A number of interesting formulas involving the group exponential map can be obtained in a finite dimensional Lie group (or more generally in a BCH-Lie group) as particular cases of the BCH formula. This formula is an infinite series of commutators that allows to write the composition of two group exponentials as a single exponential: $\exp(x)\exp(y) = \exp(BCH(x, y))$. Intuitively, this formula shows how much $\log(\exp(x)\exp(y))$ deviates from $x + y$ due to the non-commutativity of the multiplication in G. Remarkably, this deviation can be expressed only in terms of Lie brackets between x and y [God82, Chap. VI].

Theorem 5.6 (Series form of the BCH formula)**.** *Let x, y be in $\mathfrak{g}$. If they are small enough, then the logarithm of the product* $\exp(x)\exp(y)$ *is well-defined and we have the following development:*

$$\begin{aligned} BCH(x,y) &= \log(\exp(x)\ \exp(y)) \\ &= x + y + \tfrac{1}{2}([x,y]) + \tfrac{1}{12}([x,[x,y]] + [y,[y,x]]) \\ &\quad + \tfrac{1}{24}([[x,[x,y]],y]) + O((\|x\| + \|y\|)^5). \end{aligned} \tag{5.5}$$

A fundamental property of this function is the following: it is not only C^∞ but also *analytic* around 0, which means that $BCH(x,y)$ (near 0) is the sum of an absolutely converging multivariate infinite series (the usual multiplication is replaced here by the Lie bracket). This implies in particular that all the (partial) derivatives of this function are also analytic. The BCH formula is probably the most important formula from a practical point of view: it is the cornerstone approximation used to efficiently update the transformation parameters in SVF-based registration algorithms of Section 5.6.2.

5.3.2. Cartan-Schouten (CCS) Connections

For each tangent vector $x \in \mathfrak{g} \simeq T_eG$, the one parameter subgroup $\gamma_x(t)$ is a curve that starts from identity with this tangent vector. In order to see if these curves could be seen as geodesics, we now investigate natural connections on Lie groups. We start with left-invariant connections verifying $\nabla_{dL_gX} dL_g Y = dL_g \nabla_X Y$ for any vector fields X and Y and any group element $g \in G$. As the connection is completely determined by its action on the sub-algebra of left-invariant vector fields, we can restrict to the Lie algebra. Let $\tilde{x} = dL\,x$ and $\tilde{y} = dL\,y$ be two left-invariant vector fields. Stating that the covariant derivative of $\tilde{y}$ along $\tilde{x}$ is left-invariant amounts to say that the field $\nabla_{\tilde{x}}\tilde{y} = dL(\nabla_{\tilde{x}}\tilde{y}|_e)$ is determined by its value at identity $\alpha(x,y) = \nabla_{\tilde{x}}\tilde{y}|_e \in \mathfrak{g}$. Conversely, each bilinear operator of the Lie algebra $\alpha : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ uniquely defines the connection at the identity and thus on all left-invariant vector fields: $\nabla^\alpha_{\tilde{x}}\tilde{y} = \tilde{\alpha}(x,y)$. The connection is then uniquely extended to all vector fields using the linearity in the first variable and the Leibniz rule.

Definition 10 (Cartan-Schouten and Bi-invariant Connections)**.** Among the left-invariant connections, the **Cartan-Schouten connections** are the ones for which geodesics going through identity are one parameter subgroups. Bi-invariant connections are both left- and right-invariant.

The definition of Cartan-Schouten connection used here is due to [Pos01, Def.

6.2 p.71]. It generalizes the three classical +, − and 0 Cartan-Schouten connections [CS26] detailed below in Theorem 5.8.

Theorem 5.7. *Cartan-Schouten connections are uniquely determined by the property* $\alpha(x, x) = 0$ *for all* $x \in \mathfrak{g}$*. Bi-invariant connections are characterized by the condition:*

$$\alpha([z, x], y) + \alpha(x, [z, y]) = [z, \alpha(x, y)] \quad \forall x, y, z \in \mathfrak{g}. \tag{5.6}$$

The one dimensional family of connections generated by $\alpha(x, y) = \lambda[x, y]$ *satisfies these two conditions. Moreover, there is a unique symmetric bi-invariant Cartan-Schouten connection called the* **Canonical Cartan-Schouten (CCS)** *connection of the Lie group (also called mean or 0-connection) defined by* $\alpha(x, y) = \frac{1}{2}[x, y]$ *for all* $x, y \in \mathfrak{g}$*, i.e.* $\nabla_{\tilde{x}}\tilde{y} = \frac{1}{2}[\tilde{x}, \tilde{y}]$ *for two left-invariant vector fields.*

Indeed, let us consider the one-parameter subgroup $\gamma_x(t)$ starting from e with initial tangent vector $x \in \mathfrak{g}$. As this is the integral curve of the left-invariant vector field $\tilde{x} = dL\, x$, its tangent vector is $\dot{\gamma}_x(t) = dL_{\gamma_x(t)}x = \tilde{x}|_{\gamma_x(t)}$. The curve is a geodesic if and only if it is auto-parallel, i.e. if $\nabla_{\dot{\gamma}_x}\dot{\gamma}_x = \nabla_{\tilde{x}}\tilde{x} = \tilde{\alpha}(x, x) = 0$. Thus, the one-parameter subgroup $\gamma_x(t)$ is a geodesic if and only if $\alpha(x, x) = 0$.

This condition implies that the operator α is skew-symmetric. However, if any skew-symmetric operator give rise to a left-invariant connection, this connection is not always right-invariant. The connection is right-invariant if $\nabla_{dR_g X} dR_g Y = dR_g \nabla_X Y$ for any vector fields X and Y and any group element g. As we have $(dR_g\tilde{x}) = \widetilde{\mathrm{Ad}(g^{-1})x}$ for any left-invariant vector field $\tilde{x} = dL\, x$, the right-invariance is equivalent to the Ad-invariance of the operator α:

$$\alpha\left(\mathrm{Ad}(g^{-1})x, \mathrm{Ad}(g^{-1})y\right) = \mathrm{Ad}(g^{-1})\, \alpha(x, y),$$

for any two vectors $x, y \in \mathfrak{g}$ and $g \in G$. We can focus on the infinitesimal version of this condition by taking the derivative at $t = 0$ with $g^{-1} = \exp(tz)$. Since $\frac{d}{dt}\mathrm{Ad}(\exp(tz))x = [z, x]$ we obtain the requested characterization of bi-invariant connections: $\alpha([z, x], y) + \alpha(x, [z, y]) = [z, \alpha(x, y)]$.

The well known one-dimensional family of connections generated by $\alpha(x, y) = \lambda[x, y]$ obviously satisfy this condition (in addition to $\alpha(x, x) = 0$). It was shown by Laquer [Laq92] that this family basically describes all the bi-invariant connections on compact simple Lie groups (the exact result is that the space of bi-invariant affine connections on G is one-dimensional) *except* for SU(d) when $d > 3$ where there is a two-dimensional family of bi-invariant affine connections.

Torsion and Curvature of Cartan-Schouten Connections

The torsion of a connection can be expressed in the basis of left-invariant vector fields: $T(\tilde{x},\tilde{y}) = \nabla_{\tilde{x}}\tilde{y} - \nabla_{\tilde{y}}\tilde{x} - [\tilde{x},\tilde{y}] = \tilde{\alpha}(x,y) - \tilde{\alpha}(y,x) - \widetilde{[x,y]}$. This is itself a left-invariant vector field characterized by its value at identity $T(x,y) = \alpha(x,y) - \alpha(y,x) - [x,y]$. Thus, the torsion of a Cartan connection is $T(x,y) = 2\alpha(x,y) - [x,y]$. In conclusion, there is a unique torsion-free Cartan connection, called the Symmetric Cartan-Schouten connection, which is characterized by $\alpha(x,y) = \frac{1}{2}[x,y]$, i.e. $\nabla_{\tilde{x}}\tilde{y} = \frac{1}{2}[\tilde{x},\tilde{y}]$.

The curvature can also be expressed in the basis of left-invariant vector fields: $R(\tilde{x},\tilde{y})\tilde{z} = \nabla_{\tilde{x}}\nabla_{\tilde{y}}\tilde{z} - \nabla_{\tilde{y}}\nabla_{\tilde{x}}\tilde{z} - \nabla_{[\tilde{x},\tilde{y}]}\tilde{z}$. It is once again left-invariant and characterized by its value in the Lie algebra:

$$R(x,y)z = \alpha(x,\alpha(y,z)) - \alpha(y,\alpha(x,z)) - \alpha([x,y],z).$$

For connections of the form $\alpha(x,y) = \lambda[x,y]$, the curvature becomes

$$R(x,y)z = \lambda^2[x,[y,z]] + \lambda^2[y,[z,x]] + \lambda[z,[x,y]] = \lambda(\lambda-1)[[x,y],z],$$

where the last equality is obtained thanks to the Jacobi identity of the Lie bracket. For $\lambda = 0$ and $\lambda = 1$, the curvature is obviously null. These two flat connections are called the left and right (or + and -) Cartan-Schouten connections. For the CCS connection (often called mean or 0-connection), the curvature is $R(x,y)z = -\frac{1}{4}[[x,y],z]$, which is generally non zero.

Theorem 5.8 (Properties of Cartan-Schouten Connections)**.** *Among the bi-invariant Cartan-Schouten connections on a Lie group, there is a natural one-parameter family of the form* $\alpha(x,y) = \lambda[x,y]$ *that comprises three canonical connections called the* 0, $+$, $-$ *(or mean, right, right) connections:*

- *The* $-$ *connection is the unique connection for which all the left-invariant vector fields are covariantly constant along any vector field, inducing a global parallelism.*
- *The* $+$ *connection is the only connection for which all the right-invariant vector fields are covariantly constant along any vector field, inducing also a global parallelism.*
- *The CCS or 0-connection is the unique torsion free Cartan-Schouten connection. Its curvature tensor is generally non-zero but it is covariantly constant. It is thus a locally symmetric space, and the symmetry* $s_g(h) = gh^{-1}g$ *turns it into a globally affine symmetric space.*

These three connections have the same geodesics (left or right translations of one-parameter subgroups) because they share the same symmetric part $\nabla_X Y + \nabla_Y X = \partial_X Y + \partial_Y X$.

Connection		$T(\tilde{x},\tilde{y})$	$R(\tilde{x},\tilde{y})\tilde{z}$
$\nabla^{\lambda}_{\tilde{x}}\tilde{y} = \lambda[\tilde{x},\tilde{y}]$	λ	$(2\lambda-1)[\tilde{x},\tilde{y}]$	$\lambda(\lambda-1)[[\tilde{x},\tilde{y}],\tilde{z}]$
$\nabla^{-}_{\tilde{x}}\tilde{y} = 0$	*0 (-)*	$-[\tilde{x},\tilde{y}]$	*0*
$\nabla^{s}_{\tilde{x}}\tilde{y} = \frac{1}{2}[\tilde{x},\tilde{y}]$	$\frac{1}{2}$ *(0)*	*0*	$-\frac{1}{4}[[\tilde{x},\tilde{y}],\tilde{z}]$
$\nabla^{+}_{\tilde{x}}\tilde{y} = [\tilde{x},\tilde{y}]$	*1 (+)*	$+[\tilde{x},\tilde{y}]$	*0*

When we only focus on geodesics, we can thus restrict to the symmetric CCS connection. However, we should be careful that the parallel transport differ for the three connections as we will see below.

5.3.3. Group geodesics, parallel transport

We call group geodesics the geodesics of the canonical Cartan-Schouten connection. We already know that the geodesics going through identity are the one-parameter subgroups (by definition of the Cartan-Schouten connections). The canonical Cartan connection being left-invariant, the curve $\gamma(t) = g\exp(tx)$ is also a geodesic. We have indeed $\dot{\gamma} = dL_g\dot{\gamma}_x$ and $\nabla_{\dot{\gamma}}\dot{\gamma} = dL_g\nabla_{\dot{\gamma}_x}\dot{\gamma}_x = 0$. As $\gamma(0) = dL_g x$, we finally obtain that:

Theorem 5.9 (Group exponential)**.** *The group geodesic starting at g with tangent vector $v \in T_gG$ is $\gamma_{(g,v)}(t) = g\exp(t\,dL_{g^{-1}}\,v)$. Thus, the (group) exponential map at point g is:*

$$\exp_g(v) = \gamma_{g,v}(1) = g\exp(dL_{g^{-1}}v).$$

As noted in Theorem 5.1, there exists for each point g of G a normal convex neighborhood (NCN) in the sense that for any couple of points (p,q) in this neighborhood, there exists a unique geodesic of the form $\exp_p(tv)$ joining them which lies completely in this neighborhood. Furthermore, a NCN $\mathcal{V}_e$ of the identity is transported by left-invariance into a NCN $g\mathcal{V}_e$ of any point $g \in G$.

Of course, we could have defined the geodesics using the right translations to obtain curves of the form $\exp(t\,dR_{g^{-1}}\,v)\,g$. In fact, those two types of group geodesic are the same and are related by the adjoint operator, as shown below. However, we should be careful that the left and right transport of the NCN at the identity lead to different NCN of a point g: $g\mathcal{V} \neq \mathcal{V}g$.

Theorem 5.10. *Let x be in $\mathfrak{g}$ and g in G. Then we have:*

$$g\exp(x) = \exp(Ad(g)x)\,g.$$

For all g in G, there exists an open neighborhood $\mathcal{W}_g$ of $e \in G$ (namely $\mathcal{W}_g = \mathcal{V}_e \cap g\mathcal{V}_e g^{-1}$ where $\mathcal{V}_e$ is any NCN of e) such that for all $m \in \mathcal{W}_g$ the quantities $\log(m)$ *and*

$\log(g\,m\,g^{-1})$ *are well-defined and are linked by the following relationship:*

$$\log(g\,m\,g^{-1}) = Ad(g)\log(m).$$

Notice that in general the NCN $\mathcal{W}_g$ depends on g unless we can find a NCN $\mathcal{V}_e$ that is stable by conjugation. These equations are simply the generalization to (abstract) Lie groups of the well-known matrix properties: $G\exp(V)\,G^{-1} = \exp(GVG^{-1})$ and $G\log(V)\,G^{-1} = \log(GVG^{-1})$.

Corollary 1 (Group exponential and log map at any point)**.** *For all g in G, there exists an open neighborhood $\mathcal{V}_g$ of g such that the local exponential and logarithmic maps of the CCS connection are well defined and the inverse of each other. Moreover, their left and right expressions are:*

$$\exp_g(v) = g\exp(dL_{g^{-1}}v) = \exp(dR_{g^{-1}}v)\,g \quad \textit{for} \quad v \in T_gG;$$

$$\log_g(x) = dL_g\log(g^{-1}x) = dR_g\log(xg^{-1}) \quad \textit{for} \quad x \in \mathcal{V}_g.$$

Parallel Transport Along Geodesics

To obtain the parallel transport, we have to integrate the infinitesimal parallel transport given by the connection along the path. Because left and right translations commute, the parallel transport for the λ-Cartan connection of $y \in \mathfrak{g}$ to $T_{\exp(x)}G$ along the one parameter subgroup $\exp(tx)$ is:

$$\Pi_e^{\exp(x)}y = dL_{\exp((1-\lambda)x)}dR_{\exp(\lambda x)}y$$

One can verify that this formula is consistent with the infinitesimal version (the λ-Cartan connection): $\frac{d}{dt}\Pi(\exp(tx))_t^0\tilde{y} = \lambda[\tilde{x},\tilde{y}]$. In particular, we observe that the parallel transport of the left (resp. right) Cartan-Schouten connection ($\lambda = 0$, resp $\lambda = 1$) is the left (resp. right) translation. Both are independent of the path: the Lie group endowed with these connections is a spaces with absolute parallelism (but with torsion). This was expected wince the curvature of these two connections vanishes.

For the canonical symmetric Cartan-Schouten connection ($\lambda = 1/2$), the parallel transport is a geometric average where we transport on the left for half of the path and on the right for the other half: $\Pi_e^{\exp(x)}y = dL_{\exp(x/2)}dR_{\exp(x/2)}y$. Parallel translation along other geodesics is obtained by left (or right) translation or this formula.

5.4. Left, Right and Bi-invariant Riemannian Metrics on a Lie Group

In the case of Lie groups, there are two natural families of left (resp right)-invariant Riemannian metrics that are determined by an inner product at the tangent space of the identity and prolonged everywhere by left (resp. right) translation: $< v, w >^L_g =< dL_{g^{-1}} v, dL_{g^{-1}} w >_e$. Conversely, a Riemannian metric is left-invariant if all left translations are isometries. Right invariant metrics are defined similarly. It is interesting to understand if and when their geodesics coincide with the group geodesics, and if not how much they differ.

5.4.1. Levi-Civita Connections of Left-Invariant Metrics

Let us define the operator ad^* as the metric adjoint of the adjoint operator ad: this is the unique bilinear operator satisfying for all vector fields $X, Y, Z \in \Gamma(G)$:

$$\langle \mathrm{ad}^*(Y, X), Z\rangle = \langle [X, Z], Y\rangle\,.$$

This allows us to rewrite Khozul formula for the Levi-Civita connection as:

$$2\,\langle \nabla_X Y, Z\rangle = \langle [X, Y] - \mathrm{ad}^*(X, Y) - \mathrm{ad}^*(Y, X), Z\rangle + \partial_X \langle Y, Z\rangle + \partial_Y \langle X, Z\rangle - \partial_Z \langle X, Y\rangle\,.$$

Without loss of generality, we restrict to left-invariant vector fields because they constitute a basis of all vector fields (with functional coefficients). Since the dot product is covariantly constant for left-invariant metrics, we left are with

$$\nabla_{\tilde{x}} \tilde{y} = \frac{1}{2}\left([\tilde{x}, \tilde{y}] - \mathrm{ad}^*(\tilde{x}, \tilde{y}) - \mathrm{ad}^*(\tilde{y}, \tilde{x})\right). \tag{5.7}$$

Denoting $\mathrm{ad}^*(x, y) = \mathrm{ad}^*(\tilde{x}, \tilde{y})|_e$ the restriction of the ad^* operator to the Lie algebra, we end up with the following bilinear form characterizing the Levi-Civita connection of a left-invariant metric in the Lie algebra:

$$\alpha^L(x, y) = \frac{1}{2}\left([x, y] - \mathrm{ad}^*(x, y) - \mathrm{ad}^*(y, x)\right).$$

Let γ_t be a curve and $x^L_t = dL_{\gamma_t^{-1}} \dot{\gamma}_t$ be the *left angular speed vectors* (i.e. the left translation of the tangent vector to the Lie algebra). Then the curve γ_t is a geodesic of the left-invariant metric if

$$\dot{x}^L_t = \alpha^L(x^L_t, x^L_t) = -\mathrm{ad}^*(x^L_t, x^L_t). \tag{5.8}$$

This remarkably simple quadratic equation in the Lie algebra is called the Euler-Poincaré equation. A very clear derivation is given in [Kol07].

Since the CCS connection is determined by $\alpha(x, y) = \frac{1}{2}[x, y]$, the geodesic of the left-invariant metric γ_t is also a geodesic of the canonical Cartan connection if $x^L_t = x$ is constant. Thus, we see that the geodesic of a left-invariant Riemannian metric going through identity with tangent vector x is a one parameter subgroups if and only if $\mathrm{ad}^*(x, x) = 0$. Such vectors are called normal elements of the Lie algebra. For other elements, the symmetric part of ad^* encodes the acceleration of one parameter subgroups with respect to the left-invariant geodesics starting with the same tangent vector.

5.4.2. Canonical Connection of Bi-invariant Metrics

Riemannian metrics which are simultaneously left- and right-invariant are called *bi-invariant*. For these special metrics, we have the very interesting result:

Theorem 5.11. *A left-invariant metric on a Lie group is bi-invariant if and only if for all $g \in G$, the adjoint operator $Ad(g)$ is an isometry of the Lie algebra $\mathfrak{g}$:*

$$\langle Ad(g)y, Ad(g)z \rangle = \langle y, z \rangle \,,$$

or equivalently if and only if for all elements $x, y, z \in \mathfrak{g}$:

$$\langle [x,y], z \rangle + \langle y, [x,z] \rangle = 0 \quad or \quad ad^*(x,y) = -ad^*(y,x). \tag{5.9}$$

Thus, the Levi-Civita connection of a bi-invariant metric is necessarily the canonical symmetric Cartan-Schouten connection characterized by $\alpha(x,y) = \frac{1}{2}[x,y]$ on the Lie algebra. Moreover, a bi-invariant metric is also invariant w.r.t. inversion. Group geodesics of G (including one-parameter subgroups) are the geodesics of such metrics.

Equation 5.9 is the infinitesimal version of the invariance of the dot product by the adjoint group: It actually specifies that the metric dual of the adjoint ad^* is skew symmetric, i.e. that the Levi-Civita connection of the metric considered is the canonical symmetric Cartan-Schouten connection of the Lie group.

The proof is given in [Ste64, Chap. V] and [Pos01, Chap. 25]. In short, any element close to identity can be written $g = \exp(tx)$, for which we have $d\mathrm{Ad}(g)y/dt = [x,y]$. Thus differentiating $\langle \mathrm{Ad}(g)y, \mathrm{Ad}(g)z \rangle = \langle y, z \rangle$ gives the left part of equation (5.9). Since this should be verified for all $x \in \mathfrak{g}$, we obtain the skew symmetry of the ad^* operator. Thus, the mean Cartan-Schouten connection is the Levi-Civita connection of any bi-invariant metric: the affine framework corresponds to the Riemannian framework when this last one is fully consistent with the group operations.

An interesting consequence is that any Lie group with a bi-invariant metric has non-negative sectional curvature. Indeed, the sectional curvature in the two-plane $\mathrm{span}(x,y)$ for $x, y \in \mathfrak{g}$ can be computed using left-invariant vector fields:

$$k(x,y) = \frac{\langle R(x,y)y, x \rangle}{\|x\|^2 \|y\|^2 - \langle x,y \rangle^2} = \frac{1}{4} \frac{\|[x,y]\|^2}{\|x\|^2 \|y\|^2 - \langle x,y \rangle^2}, \tag{5.10}$$

where we used the expression $R(x,y)z = -\frac{1}{4}[[x,y],z]$ of the Riemannian curvature and Eq. 5.9 to move one bracket from left to right in the inner product. Thus, taking two orthonormal vectors of the Lie algebra, the section curvature reduces to $k(x,y) = \frac{1}{4}\|[x,y]\|^2$ which is non-negative.

Compactness, Commutativity and Existence of Bi-invariant Metrics

The next question is to understand under which conditions such a bi-invariant metric exists. From Theorem 5.11, we see that if a bi-invariant metric exists for the Lie group, then $\mathrm{Ad}(g)$ is an isometry of $\mathfrak{g}$ and can thus be looked upon as an element of the orthogonal group $\mathrm{O}(d)$ where $d = \dim(G)$. As $\mathrm{O}(d)$ is a *compact* group, the adjoint group $\mathrm{Ad}(G) = \{\mathrm{Ad}(g)/g \in G\}$ is necessarily *included in a compact set*, a situation called *relative compactness*. This notion actually provides a sharp criterion, since the theory of differential forms and their integration can be used to explicitly construct a bi-invariant metric on relatively compact subgroups [Ste64, Theorem V.5.3.].

Theorem 5.12. *The Lie group G admits a bi-invariant metric if and only if its adjoint group $Ad(G)$ is relatively compact.*

In the case of *compact* Lie groups, the adjoint group is also compact, and Theorem 5.12 implies that bi-invariant metrics exist. This is the case of *rotations*, for which bi-invariant Fréchet means have been extensively studied and used in practical applications, for instance in [Pen96, Pen98, Moa02]. In the case of commutative Lie groups, left and right translations are identical and any left-invariant metric is trivially bi-invariant. Direct products of compact Abelian groups obviously admit bi-invariant metrics but Theorem 5.12 shows that in the general case, non-compact and non-commutative Lie groups which are not the direct product of such groups may fail to admit a bi-invariant metric.

Bi-invariant Pseudo-Riemannian Metrics (Quadratic Lie Groups)

The class of Lie groups that admits bi-invariant metrics is quite small and does not even include rigid body transformations in 2D or 3D. Looking for a bi-invariant Riemannian framework for larger groups is thus hopeless. One possibility is to relax the positivity of the Riemannian metric to focus on bi-invariant *pseudo-Riemannian* metrics only, for which the bilinear forms on the tangent space of the manifold are definite but may not be positive (eigenvalues are just non-zero). Based on the classification of quadratic Lie algebras of [Med82, MR85], Miolane developed in [MP15] an algorithm to compute all the bi-invariant pseudo-metric on a given Lie group (if they exist). This algorithm was applied to simple Lie groups that have a locally unique bi-invariant mean (scaling and translations, the Heisenberg group, upper triangular matrices and Euclidean motions). It showed that most of them do not possess a bi-invariant pseudo-metric. The special Euclidean motion group in 3D is a notable exception. Thus, the generalization of the statistical theory to pseudo-Riemannian metrics may not be so interesting in practice.

5.4.3. Example with Rigid Body Transformations

In medical imaging, the simplest possible registration procedure between two images uses rigid-body transformations, characterized by a rotation matrix and a translation vector. Since there exist bi-invariant metrics on rotations and on translations, one could hope for the existence of bi-invariant metrics. We show below that this is not the case.

The Lie group of rigid-body transformations in the d-dimensional Euclidean space, written here $\mathrm{SE}(d)$, is the semi-direct product of $(\mathrm{SO}(d), \times)$ (rotations) and $(\mathbb{R}^d, +)$ (translations). An element of $\mathrm{SE}(d)$ is uniquely represented by a couple $(R, t) \in \mathrm{SO}(d) \ltimes \mathbb{R}^d$ with the action on a point x of $\mathbb{R}^d$ being defined by $(R, t).x = Rx + t$. The multiplication is then $(R', t')\,(R, t) = (R'R, R't + t')$, the neutral element $(\,\mathrm{Id}, 0)$ and the inverse $(R^\top, -R^\top t)$. The fact that the product between rotations and translations is semi-direct and not direct (there is a coupling between rotation and translation in the composition) is at the heart of the non-existence of a bi-invariant metric on the product group.

We obtain a faithful representation of $\mathrm{SE}(d)$ and its Lie algebra using homogeneous coordinates: $(R, t) \simeq \left(\begin{smallmatrix} R & t \\ 0 & 1 \end{smallmatrix}\right)$ and $(\Omega, v) \simeq \left(\begin{smallmatrix} \Omega & v \\ 0 & 0 \end{smallmatrix}\right)$, where Ω is any skew $d \times d$ matrix and v any vector of $\mathbb{R}^d$. In the homogeneous representation, the Lie bracket $[.,.]$ is simply the matrix commutator, which gives the following Lie bracket for the Lie algebra $\mathfrak{se}(n) = \mathfrak{so}(d) \ltimes \mathbb{R}^d$: $[(\Omega, v), (\Omega', v')] = (\Omega\Omega' - \Omega'\Omega,\ \Omega v' - \Omega' v)$.

Proposition 1. *The action of the adjoint operator Ad of the group of rigid-body transformations* $\mathrm{SE}(d)$ *at the point* (R, t) *on an infinitesimal displacement* $(\Omega, v) \in \mathfrak{se}(d)$ *is given by:*

$$Ad(R, t)(\Omega, v) = (R\Omega R^\top,\ -R\Omega R^\top t + Rv).$$

As it is unbounded, no bi-invariant Riemannian metric exists on the space of rigid-body transformations for $d > 1$.

Such a result was already known for SE(3) [ZKC99]. It is established here for all dimensions. A very interesting result of [MP15] is that there is no bi-invariant pseudo-Riemannian metric either on $\mathrm{SE}(d)$ either, except for $d = 3$.

5.5. Statistics on Lie Groups as Symmetric Spaces

Although bi-invariant metrics may fail to exist, any Lie group has a symmetric canonical Cartan-Schouten connection for which group means can be defined implicitly as exponential barycenters, at least locally. As will be shown in the sequel, this definition has all the desirable bi-invariance properties, even when bi-invariant metrics do not exist. Moreover, we can show the existence and uniqueness of the group mean even globally in a number of cases. Group means were investigated for compact and nilpotent Lie groups in [BK81, Chapter 8] as a side results in the study of almost flat metrics. In the medical image analysis domain, group means were originally proposed in [APA06] under the name “bi-invariant means” and fully developed as exponential barycenters of the CCS connection in [PA12].

5.5.1. Bi-invariant Means with Exponential Barycenters of the CCS Connection

Every Lie group is orientable and has a Haar volume form. When the group is unimodular (for instance in SE(d)), the Haar measure is bi-invariant and provides a canonical reference measure to define intrinsic probability density functions on the group. However, left and right Haar measures differ by a function (the determinant of the adjoint) in non-unimodular groups such as GL(d) so that this property cannot be used in general. In any case, we can consider probability measures on the group $\mu \in Prob(G)$ that include sample distributions $\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{g_i}$ of n transformations g_i in the group G. The left, right translations and inversion being smooth and one-to-one, there is no problem to consider their push-forward action on measures, which write in the case of the above sample distribution: $dL_h\mu = \sum_{i=1}^{n}\delta_{hg_i}$, $dR_h\mu = \sum_{i=1}^{n}\delta_{g_ih}$ and $d\mathrm{Inv}\mu = \sum_{i=1}^{n}\delta_{g_i^{-1}}$.

Definition 11 (Groups means)**.** Let $\mu \in Prob(G)$ be a probability distribution on G with compact support Supp(μ) belonging to an open set $\mathcal{V}$ diffeomorphic to a sphere such that $\log(g^{-1}h)$ and $\log(hg^{-1}) = \mathrm{Ad}(g)\log(g^{-1}h)$ exists for any point $g \in \mathcal{V}$ and $h \in \mathrm{Supp}(\mu)$. The points $\bar{g} \in \mathcal{V}$ solutions of the following group barycentric equation (if there are some) are called group means:

$$\int_G \log(\bar{g}^{-1}h)\,\mu(dh) = 0. \tag{5.11}$$

This definition translates the exponential barycenters in the language of Lie groups. However, the definition is a bit more general since we do not require $\mathcal{V}$ to be a convex normal neighborhood. Therefore, we cannot conclude about existence in general but we know that it is ensured for a sufficiently concentrated distribution. This definition is close to the Riemannian center of mass (or more specifically the Riemannian average) of [Gro04] but uses the group logarithm instead of the Riemannian logarithm. Notice that the group geodesics cannot be seen as Riemannian geodesics when the CCS connection is non-metric so that the group means are different (in general) from the Fréchet or Karcher mean of some Riemannian metric.

The consistency constrain under the adjoint action on this neighborhood was also added to obtain the following equivariance theorem, which is stated with empirical means (sample distributions) for simplicity.

Theorem 5.13 (Equivariance of group means)**.** *Group means, when they exist, are left-*

, right- and inverse-equivariant: if $\bar{g}$ is a mean of n points $\{g_i\}$ of the group and $h \in G$ is any group element, then $h\bar{g}$ is a mean of the points hg_i, $\bar{g}h$ is a mean of the points $\{g_i h\}$ and $\bar{g}^{-1}$ is a mean of $\{g_i^{-1}\}$.

Proof. We start with the left-equivariance. If $\bar{g}$ is a mean of the points $\{x_i\}$ and $h \in G$ is any group element, then $\log((h\bar{g})^{-1}hg_i) = \log(g^{-1}g_i)$ exists for all points hg_i. Thus the point $h\bar{g} \in h\mathcal{V}$ is a solution of the barycentric equation $\sum_i \log((h\bar{g})^{-1}hg_i) = 0$, which proves the left-equivariance. For the right-invariance, we have to apply Theorem 5.10: $\mathrm{Ad}(\bar{g})\left(\sum_i \log(\bar{g}^{-1}g_i)\right) = \sum_i \log(g_i\bar{g}^{-1})$. Since $\mathrm{Ad}(\bar{g})$ is invertible, the usual barycentric equation, which is left-invariant, is equivalent to a right-invariant barycentric equation, and $\bar{g}h$ is a mean of the points $\{g_i h\}$. The equivariance with respect to inversion is obtained with $\sum_i \log(g_i^{-1}\bar{g}) = -\sum_i \log(\bar{g}^{-1}g_i)$. □

Theorem 5.14 (Local uniqueness and convergence to the group mean)**.** *If the transformations $\{g_i\}$ belong to a sufficiently small normal convex neighborhood $\mathcal{V}$ of some point $g \in G$, then there exists a unique solution of Eq. (5.11) in $\mathcal{V}$. Moreover, the following iterated fixed point strategy converges at least at a linear rate to this unique solution:*

1. *Initialize $\bar{g}_0$, for example with $\bar{g}_0 := g_1$.*
2. *Iteratively update the estimate of the mean:*

$$\bar{g}_{t+1} := \bar{g}_t \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log(\bar{g}_t^{-1}g_i)\right). \tag{5.12}$$

Until convergence: $\|\log(\bar{g}_t^{-1}\bar{g}_{t+1})\| < \epsilon \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\log(\bar{g}_t^{-1}g_i)\right\|^2}$.

The proof is detailed in [PA12]. The uniqueness of the mean and the convergence rate are linked to the contraction properties of the above iteration. The proof rely on an auxiliary metric $\|.\|$ on $\mathfrak{g}$ such that for all x, y sufficiently small, we have: $\|[x,y]\| \leq \|x\|\,\|y\|$. The left-invariant function $\phi(g,h) = \|\log(g^{-1}h)\|^2$ can then be used as a surrogate of a distance in this neighborhood. This function is sometimes called the group distance. However, this is an improper name since the triangular inequality is generally not respected. Notice that this is not a left-invariant Riemannian metric either because we use the group logarithm and not the left-invariant Riemannian log.

As in the case of the Fréchet mean, there is a closed form for the group mean of two points since this point is on the geodesic joining them.

Proposition 2. *Let h be in a normal convex neighborhood of $g \in G$. Then the group* *mean of g and h (with weights $1-\alpha$ and $\alpha > 0$) is given by:*

$$\bar{g}_\alpha = g \exp\left(\alpha \log(g^{-1}h)\right) = g(g^{-1}h)^\alpha. \tag{5.13}$$

This explicit formula is quite exceptional. In general, there is no closed form for the group mean of more than 2 points. However, there are some specific groups where a closed form exists for the bi-invariant mean in all cases as we will detail in next section.

5.5.2. Existence and Uniqueness Results in Specific Matrix Groups

First it is worth noticing that compact Lie groups carry a bi-invariant metric for which we can use the Riemannian result. Using a normalized metric in $\mathfrak{g}$ such that $\|\mathrm{ad}x\| \leq \|x\|$, [BK81] established the convexity of geodesic balls of radius $\rho < \pi/2$ so that the group mean is unique for distributions with support in a geodesic ball of radius strictly less than $\pi/4$. [BK81] also considered simply connected nilpotent Lie groups endowed with the canonical symmetric Cartan-Schouten connection. Despite the fact that these groups generally cannot carry bi-invariant metrics (their Killing form is zero) they proved that any distribution with compact support has a unique group mean. We summarize their results in the following theorem without demonstration

Theorem 5.15 (Uniqueness of group means in compact Lie groups [BK81])**.** *A compact connected Lie group can be endowed with a bi-invariant metric normalized such that $\|adx\| \leq \|x\|$, which is compatible with the canonical symmetric Cartan-Schouten connection. With these conventions, any probability distribution with support in a regular geodesic ball of radius strictly less than $\pi/4$ has a unique group mean.*

Theorem 5.16 (Uniqueness of group means in nilpotent Lie groups [BK81])**.** *A probability with compact support in a simply connected nilpotent Lie group endowed with the canonical symmetric Cartan-Schouten connection has a unique group mean.*

The Heisenberg group below is a good example of a simply connected nilpotent Lie group that has no bi-invariant metric but a unique group mean. Then we turn to the more general class of solvable Lie groups which was not addressed by Buser and Karcher. With scaling-translations $\mathrm{ST}(d)$ and the upper triangular matrices with scalar diagonal $\mathrm{UT}(d)$, we give two examples of non-nilpotent groups with no bi-invariant metric that have unique group means as well. Moreover, the computation of the group mean can be done in finite time using an iterative scheme solving each coordinate at a time. We conjecture that this could be a feature of solvable groups. Last but not least, we look at rigid body transformations and some properties of $\mathrm{GL}(d)$. These example are summarized below, we refer the reader to [PA12] for more details.

5.5.2.1. The Heisenberg Group

This is the group of 3D upper triangular matrices of the form: $\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix}$. Parameterizing each element by the triplet (x, y, z), the group composition is $(x_1, y_1, z_1)(x_2, y_2, z_2) = (x_1 + x_2, y_1 + y_2, z_1 + z_2 + x_1 y_2)$. The Heisenberg group is thus a semi-direct product of $(\mathbb{R}^2, +)$ and $(\mathbb{R}, +)$, which is not commutative. The inversion is $(x, y, z)^{-1} = (-x, -y, -z + xy)$ with neutral element $(0, 0, 0)$.

The entire Heisenberg group is a normal convex neighborhood and we have:

$$\exp(u, v, w) = (u, v, w + \tfrac{1}{2}uv),$$
$$\log(x, y, z) = (x, y, z - \tfrac{1}{2}xy).$$

Proposition 3. *The action of the adjoint operator Ad of the Heisenberg group at a point (x, y, z) on an infinitesimal displacement (u, v, w) is given by:*

$$Ad(x, y, z)(u, v, w) = (u, v, -y\,u + x\,v + w).$$

Because it is unbounded, no bi-invariant metric exists. However, the group mean $(\bar{x}, \bar{y}, \bar{z})$ of a set of points $\{(x_i, y_i, z_i)\}_{1 \le i \le n}$ in the Heisenberg group is unique and given explicitly by:

$$(\bar{x}, \bar{y}, \bar{z}) = \frac{1}{n}\left(\textstyle\sum_i x_i, \sum_i y_i, \sum_i z_i + \frac{1}{2}\left(\left(\sum_i x_i\right).\left(\sum_i y_i\right) - \sum_i x_i y_i\right)\right).$$

5.5.2.2. Scaling and Translations ST(d)

The group of scaling and translations in dimension d is one of the simple cases of non-compact, non-commutative, non-nilpotent but solvable Lie groups which does not possess any bi-invariant Riemannian metric. An element of $\mathrm{ST}(d)$ can be uniquely represented by a scaling factor $\lambda \in \mathbb{R}_+^\star$ and a a translation $t \in \mathbb{R}^d$. The action of an element $(\lambda, t) \in \mathbb{R}_+^\star \ltimes \mathbb{R}^d$ on a vector $x \in \mathbb{R}^d$ is: $(\lambda, t)x = \lambda x + t$. Accordingly, the composition in $\mathrm{ST}(d)$ is: $(\lambda', t')\,(\lambda, t) = (\lambda'\lambda,\ \lambda' t + t')$ and inversion is $(\lambda, t)^{-1} = (1/\lambda, -t/\lambda)$. A faithful matrix representation of $\mathrm{ST}(d)$ is the subgroup of triangular matrices of the form $\begin{pmatrix} \lambda\, \mathrm{Id}_d & t \\ 0 & 1 \end{pmatrix}$. The elements of the Lie algebra are of the form (α, v) where $\alpha \in \mathbb{R}$ and $v \in \mathbb{R}^d$. Using e^α and $\ln(\lambda)$ to denote the scalar exponential and logarithm functions, the group exponential and logarithm can be written:

$$\exp(\alpha, v) = (e^\alpha, (e^\alpha - 1)/\alpha\, v)\,,$$
$$\log(\lambda, t) = (\ln(\lambda), \ln(\lambda)/(1 - \lambda)\, t)\,.$$

Once again, the entire space $\mathrm{ST}(d)$ is a normal convex neighborhood: any two points can be joined by a unique group geodesic.

The adjoint action on the Lie algebra can be computed in the matrix representation:

$$\mathrm{Ad}((\lambda,t))(\alpha,v) = \begin{pmatrix} \lambda\ \mathrm{Id}_d & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha\ \mathrm{Id}_d & v \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda}\ \mathrm{Id}_d & -\frac{t}{\lambda} \\ 0 & 1 \end{pmatrix}$$

Proposition 4. *The action of the adjoint operator Ad of the* $\mathrm{ST}(d)$ *group at a point* (λ, t) *on an infinitesimal displacement* (α, v) *is given by:*

$$Ad((\lambda,t))(\alpha,v) = (\alpha,\ \lambda v - \alpha t).$$

Because the translation t and the scale λ *in* $\lambda v - \alpha t$ *are unbounded, no bi-invariant metric exists on* $\mathrm{ST}(d)$*, although it is the semi-direct product of two commutative groups. Nevertheless, the group mean* $(\bar{\lambda}, \bar{t})$ *of a set of points* $\{(\lambda_i, t_i)\}_{1\le i\le n}$ *in the* $\mathrm{ST}(d)$ *group is unique and given explicitly by:*

$$\bar{\lambda} = \exp\left(\tfrac{1}{n}\textstyle\sum_i \ln(\lambda_i)\right), \quad \bar{t} = \frac{\sum_i \alpha_i\, t_i}{\sum_i \alpha_i}, \quad with \quad \alpha_i = \frac{\ln\left(\lambda_i/\bar{\lambda}\right)}{\lambda_i/\bar{\lambda} - 1}. \qquad (5.14)$$

5.5.2.3. Scaled Upper Uni-triangular Matrix Group

We consider the group $\mathrm{UT}(d)$ of $d \times d$ upper triangular with scaled unit diagonal. Such matrices have the form $M = \lambda\ \mathrm{Id} + N$, where λ is a positive scalar, Id the identity matrix and N an upper triangular nilpotent matrix ($N^d = 0$) with only zeros in its diagonal. This group generalizes the Heisenberg group, which is the subgroup of matrices of $\mathrm{UT}(3)$ with $\lambda = 1$. The group composition is: $M'M = (\lambda'\,\lambda)\,\mathrm{Id} + (\lambda'\,N + \lambda\,N' + N'N)$. The nilpotency of N allows to write the inversion in closed form:

$$M^{-1} = (\lambda\ \mathrm{Id} + N)^{-1} = \lambda^{-1}\ \left(\mathrm{Id} + \tfrac{N}{\lambda}\right)^{-1} = \lambda^{-1}\ \textstyle\sum_{k=0}^{n-1}(-1)^k\ \frac{N^k}{\lambda^k}.$$

The group exponential and logarithm are:

$$\exp(X) = \exp(\mu\,\mathrm{Id} + Y) = \exp(\mu\,\mathrm{Id})\exp(Y) = e^{\mu}\textstyle\sum_{k=0}^{n-1}\frac{(Y)^k}{k!},$$
$$\log(M) = \log\left((\lambda\, Id)\left(\mathrm{Id} + \tfrac{1}{\lambda}\,N\right)\right) = \ln(\lambda)\ \mathrm{Id} + \textstyle\sum_{k=1}^{n-1}\frac{(-1)^{k+1}}{k}\frac{N^k}{\lambda^k}.$$

Using these closed forms, one can derive the following equation:

$$\log(M'M) = \ln(\lambda'\lambda)\ \mathrm{Id} + \sum_{k=1}^{n-1}\frac{(-1)^{k+1}}{k}\left(\frac{1}{\lambda}\,N + \frac{1}{\lambda'}\,N' + \frac{1}{\lambda'\lambda}NN'\right)^k,$$

which in turn allows to simplify the equation $\sum_i \log(\bar{M}^{-1}M_i) = 0$ satisfied by the group mean $\bar{M} = \bar{\lambda}\ \mathrm{Id} + \bar{N}$ in $\mathrm{UT}(d)$. Using $\sum_i \log(\bar{M}^{-1}M_i) = -\sum_i \log(M_i^{-1}\bar{M})$,

$$\textstyle\sum_i\left(\ln(\bar{\lambda}{\lambda_i}^{-1})\ \mathrm{Id}\right) + \sum_i\sum_{k=1}^{n-1}\frac{(-1)^{k+1}}{k}\left(\frac{1}{{\lambda_i}^{-1}}{N_i}^{-1} + \frac{1}{\bar{\lambda}}\bar{N} + \frac{1}{\bar{\lambda}{\lambda_i}^{-1}}{N_i}^{-1}\bar{N}\right)^k = 0, \qquad (5.15)$$

where ${N_i}^{-1}$ is the nilpotent part of ${M_i}^{-1}$. To solve this equation, we see that $\bar{\lambda}$ is the geometric mean of the λ_i, and that the coefficient of $\bar{N}$ can be recursively computed, starting from coefficients above the diagonal. The key idea is that the k^{th} power of a nilpotent matrix N will have non-zero coefficients only in its k^{th} upper diagonal. Thus, we only need to consider the terms ${N_i}^{-1}/{\lambda_i}^{-1} + \bar{N}/\bar{\lambda} = 0$ to compute the coefficients of $\bar{M}$ above the diagonal. These coefficients are uniquely defined as a weighted arithmetic mean of the coefficients above the diagonal in the data. Thanks to this result, we can compute the above secondary diagonal elements which will be a weighted arithmetic mean of the corresponding coefficients in the data with a quadratic correction involving the previous coefficients. One can continue this way until all the coefficients of the mean have been effectively computed.

Proposition 5. *The adjoint group of the group* $\mathrm{UT}(d)$ *of* $d \times d$ *upper triangular with scaled unit diagonal includes in particular the adjoint group of the subgroup of scaling and translations* $\mathrm{ST}(d)$. *Thus, it is unbounded and there exists no bi-invariance metric. Nevertheless, the group mean* $(\bar{\lambda}\ \mathrm{Id} + \bar{N})$ *of a set of points* $\{(\lambda_i\ \mathrm{Id} + N_i)\}_{1 \leq i \leq n}$ *in the* $\mathrm{UT}(d)$ *group is unique and can be computed by a sequence of* d *geometric means on the diagonal followed by* $d-1, d-2, \ldots 1$ *arithmetic means (with a polynomial of degree* $1, 2, \ldots d-1$ *correction based on the already computed coefficients) for the coefficient of each parallel above the diagonal.*

5.5.2.4. General Rigid-Body Transformations

We have seen in Section 5.4.3 that no bi-invariant metric exists in the rigid-body case. One may now ask the question: is there a simple criterion for the existence/uniqueness of the bi-invariant group mean of rigid-body transformations? We use in this section the notations previously introduced in Section 5.4.3. Recall that $(R, t) \in \mathrm{SE}(d) = \mathrm{SO}(d) \ltimes \mathbb{R}^d$ can be faithfully represented by the homogeneous matrix embedding $\left(\begin{smallmatrix} R & t \\ 0 & 1 \end{smallmatrix}\right)$, and the Lie algebra by matrices $(\Omega, v) \simeq \left(\begin{smallmatrix} \Omega & v \\ 0 & 0 \end{smallmatrix}\right)$, where Ω is a skew $d \times d$ matrix and v a vector of $\mathbb{R}^d$. The group exponential can be computed using directly the matrix representation, or by identifying the one-parameter subgroups of $\mathrm{SE}(d)$:

$$\exp(\Omega, v) = (\exp(\Omega),\ M(\Omega)\, v) \quad \text{with} \quad M(\Omega) = \sum_{k=0}^{\infty} \frac{\Omega^k}{(k+1)!} = \exp(\Omega) \int_0^1 \exp(-t\Omega) dt.$$

To compute explicitly the value of the matrix $M(\Omega)$, we use the fact that any skew symmetric matrix can be diagonalized in the block diagonal form $U \Omega U^\top = \mathrm{diag}\left(\{\theta_j J\},\ 0\right)$ where U is an orthonormal matrix, θ_i are the 2D rotation angles and $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ is a normalized 2D skew-symmetric matrix (see Chapter 1 Section 1.7.3 on rotations). In this coordinate system, we have already computed in Chapter 1 that $\exp(\theta_j J)$ is a 2D rotation matrix of angle θ_j. A few extra steps yield:

$$M(\theta_j\, J) = \frac{\sin\theta_j}{\theta_j}\,\mathrm{Id}_2 + \frac{\cos\theta_j - 1}{\theta_j} J = \frac{1}{\theta_j}\begin{pmatrix} \sin\theta_j & \cos\theta_j - 1 \\ -\cos\theta_j + 1 & \sin\theta_j \end{pmatrix}.$$

The determinant of this matrix is $\det(M(\theta_j\, J)) = 2(1-\cos\theta_j)/\theta_j^2 > 0$ for $0 < |\theta_j| < 2\pi$, so that $M(\Omega)$ is invertible in this domain. A direct computation shows that the inverse of $M(\theta_j\, J)$ can be written:

$$M(\theta_j\, J)^{-1} = \frac{\theta_j \sin\theta_j}{2(1-\cos\theta_j)}\,\mathrm{Id}_2 + \frac{\theta_j}{2} J = \begin{pmatrix} \frac{\theta_j\sin\theta_j}{2(1-\cos\theta_j)} & \frac{\theta_j}{2} \\ -\frac{\theta_j}{2} & \frac{\theta_j\sin\theta_j}{2(1-\cos\theta_j)} \end{pmatrix}.$$

Proposition 6. *The principal logarithm of a rigid-body transformation* (R, t) *is well-defined if and only if the logarithm of its rotation part* R *is well-defined, i.e. if the angles of the 2D rotations of its decomposition are strictly less than* π *in absolute value. In that case* $M(\Omega)$ *is invertible and we get the group logarithm on* $\mathrm{SE}(d)$ *with:*

$$\log(R, t) = (\Omega, v) \quad \textit{with} \quad \Omega = \log(R),\;\; v = M(\Omega)^{-1} t \quad \textit{or} \quad \begin{pmatrix} \Omega & v \\ 0 & 0 \end{pmatrix} = \log\begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}$$

The proof rely on the block upper-triangular structure of the homogeneous matrix representing the transformation (R, t): the eigenvalues of such a matrix depend only on the blocks in its diagonal, i.e. only on R and 1 here, and not on t.

This theorem highlight the fact that the difficult part is the rotation group. Because the symmetric Cartan-Schouten connection on rotations is the Levi-Civita connection of the bi-invariant metric developed in Chapter 1 (Section **??**), we know by Karcher theorem (Chapter 2, Section **??**) that the mean rotation is unique if the support of the rotation distribution is contained in a geodesic ball of radius $r < r^* = \pi/2$). Remarkably, this is sufficient to guarantee the existence and uniqueness of the bi-invariant mean of rigid-body transformations.

Proposition 7. *Let* $\{R_i, t_i\}$ *be a set of rigid-body transformations with rotations within a geodesic ball* $\mathcal{B}_r$ *of radius* $r < \frac{\pi}{2}$*. Then the bi-invariant Riemannian mean* $\bar{R}$ *of their rotation parts is unique in this geodesic ball and there exists a unique group mean* $(\bar{R}, \bar{t})$ *on* $\mathrm{SE}(d)$ *with* $\bar{R} \in \mathcal{B}_r$*.*

Proof. We are looking for the solutions $(\bar{R}, \bar{t})$ of the group barycentric equation:

$$\textstyle\sum_i \log((\bar{R}, \bar{t})^{-1}(R_i, t_i)) = \sum_i \log\left(\bar{R}^{-1} R_i,\; \bar{R}^{-1}(t_i - \bar{t})\right).$$

Denoting $\Omega_i = \log(\bar{R}^{-1} R_i)$, this boils down to $\sum_i \Omega_i = 0$ for the rotation part, which is uniquely satisfied by the bi-invariant rotation mean $\bar{R}$ in the geodesic ball $\mathcal{B}_r$ by assumption. For the translation part, we get: $\sum_i M(\Omega_i)^{-1}\bar{R}^{-1}(t_i - \bar{t}) = 0$. Thus, if $M = \sum_i M(\Omega_i)^{-1}$ is invertible, then this equation has a unique solution:

$$\bar{t} = \bar{R}\left(M^{-1}\right)^{-1} \sum_i M(\Omega_i)^{-1}\bar{R}^{-1}t_i.$$

To show that this is the case under our assumptions, we note that $M(\Omega)^{-1} = S + \Omega/2$, where S is a symmetric matrix with positive eigenvalues $\frac{\theta_j \sin\theta_j}{2(1-\cos\theta_j)} > 0$ whenever the 2D rotation angles satisfy $|\theta_j| < \pi$. This is a consequence of the bloc diagonalization $UM(\Omega)^{-1}U^\top = \mathrm{diag}\left(\{M(\theta_j J)^{-1}\},\ 1\right)$ with $M(\theta_j\, J)^{-1} = \frac{\theta_j \sin\theta_j}{2(1-\cos\theta_j)}\,\mathrm{Id}_2 + \frac{\theta_j}{2}J$. Thus, the sum $M = \sum_i M(\Omega_i)^{-1} = \tilde{S} + \tilde{A}$ of the matrices of this type can also be written as the sum of a positive definite symmetric matrix $\tilde{S} = \sum_i S_i$ (the convex combination of SPD matrices is a SPD matrix) and a skew symmetric matrix $\tilde{A} = \sum_i \Omega_i/2$. It is thus invertible because $(\tilde{S} + \tilde{A})x = 0$ implies $x^\top \tilde{S} x = 0$ (the skew symmetry of $\tilde{A}$ implies $x^\top \tilde{A} x = 0$) which is zero only for $x = 0$ by the definiteness of $\tilde{S}$. □

Example with 2D Rigid Transformations

A 2D rigid transformation can be parametrized by $T = (\theta, t_1, t_2)$, where $\theta \in]-\pi, \pi]$ is the angle of the rotation of $SO(2) \simeq S_1$ and $t = [t_1, t_2] \in \mathbb{R}^2$ is the translation vector. We consider the following example of three rigid transformations proposed in [Pen06b, p.31] to show that left and right invariant Riemannian means were different:

- $T_1 = (\pi/4\,,\ -\sqrt{2}/2,\ \sqrt{2}/2)$,
- $T_2 = (0\,,\ \sqrt{2},\ 0)$,
- $T_3 = (-\pi/4\,,\ -\sqrt{2}/2,\ -\sqrt{2}/2)$.

A left-invariant Fréchet mean can also be computed explicitly in this case thanks to the simple form taken by the corresponding geodesics. The analogous right-invariant Fréchet mean can be computed by inverting the data, computing their left-invariant mean and then inverting this Fréchet mean. The bi-invariant mean can be computed with the above formula. This yields (after a number of simple but tedious algebraic manipulations):

- Left-invariant Fréchet mean: $(0\,,\ 0,\ 0)$,
- Bi-invariant mean: $(0\,,\ (\sqrt{2} - \frac{\pi}{4})/(1 + \frac{\pi}{4}.(\sqrt{2}+1)),\ 0) \simeq (0\,,\ 0.2171,\ 0)$,
- Right-invariant Fréchet mean: $(0\,,\ \sqrt{2}/3,\ 0) \simeq (0\,,\ 0.4714,\ 0)$.

As expected, the mean rotation angle is exactly the same in all cases. But the mean translations are different, and the bi-invariant mean is located nicely *between* the left- and right-invariant Fréchet means. This is quite intuitive, since the bi-invariant mean can be looked upon as an in-between alternative with regard to left- and right-invariant Fréchet means.

5.6. The Stationary Velocity Fields (SVF) Framework for Diffeomorphisms

Non-linear registration aims at maximizing the geometrical similarity of anatomical images by optimizing the spatial deformations that are acting on them. In this process, deformation fields quantify the anatomical changes as local changes of coordinates. Thus, deformations represent a powerful and rich geometrical object for the statistical analysis of motion and differences across organs.

In the context of medical image registration, *diffeomorphic registration* restricts the set of possible spatial transformations to diffeomorphisms. There is a rich mathematical background for the estimation and analysis of deformations that brings key properties to the analysis of medical images. Diffeomorphic registration was introduced with the "Large Deformation Diffeomorphic Metric Mapping (LDDMM)" framework [Tro98, BMMTY05], which parametrizes deformations with the flow of *time varying velocity fields* $v(x,t)$ with a right-invariant Riemannian metric, as described in Chapter 4. In view of reducing computational and memory costs of LDDMM, [ACPA06] subsequently proposed to restrict this parametrization to the subset of diffeomorphisms parametrized by the flow of *stationary velocity fields* (SVF). This setting is of particular methodological interest, since the flow associated to a SVF is a one-parameter subgroup, and we can thus take advantage of the properties of the associated Lie algebra to develop efficient and flexible computational tools. In the sequel, we review the main aspects of SVF-based image registration, with a particular focus on the computational schemes arising from the one-parameter subgroup properties.

5.6.1. Parameterizing Diffeomorphisms with the Flow of SVFs

Before investigating in depth the computational aspects of SVF-based registration, we illustrate in this section the theoretical background and rationale of the SVF framework. In order to work in a well-posed space of deformations, we need to specify the space on which the Lie algebra is modeled. This is the role of the regularization term of the SVF registration algorithms [VPPA08, HBO09] or of the spline parametrization of the SVF in [Ash07, MRD+11]: this regularization restricts the Lie algebra to sufficiently regular velocity fields. One usually chooses a sufficiently powerful Hilbert metric to model the Lie Algebra. The flow of these stationary velocity fields and their finite composition generates a subgroup of all diffeomorphisms which is the group that we consider. Up to now, the theoretical framework is very similar to the LDDMM setting: if we model the Lie algebra on the same admissible Hilbert space than LDDMM, then all the diffeomorphisms generated by the one-parameter subgroups (the exponential of SVFs) also belong to the LDDMM group. As in finite dimension, the affine geodesics of the CCS connection (group geodesics) are metric-free. Thus, these group geodesics generally differ from the Riemannian geodesics of LDDMM.

However, it is well known that the above construction generates a group which is significantly larger than the space covered by single group exponentials. Although our affine connection space is geodesically complete (all geodesics can be continued for all time without hitting a boundary), there is no Hopf-Rinow theorem in affine geometry that can ensure that any two points can be joined by a geodesic. Thus, in general, not all the elements of the group may be reached by a one-parameter subgroup. An example in finite dimension that we have already seen is $SL(2,\mathbb{R})$ (2-dimensional square matrices with unit determinant) where elements with trace less than -2 cannot be reached by any one-parameter subgroup. In the image registration context, this is generally not a problem since all the possible diffeomorphisms of the group are not as likely and we are only interested in admissible anatomical transformations. Another potential problem is that, contrarily to the finite dimension, the exponential map is not in general a diffeomorphism from a neighborhood of zero in the Lie algebra onto a neighborhood of the identity in the infinite dimensional group. For instance, there exist diffeomorphisms in every neighborhood of the identity in $\mathrm{Diff}^s(\mathcal{M})$ which are not the exponential of an H^s vector field. A classical example of the non-surjectivity of the exponential map is the function $f_{n,\epsilon}(\theta) = \theta + \pi/n + \epsilon \sin^2(n\theta)$ in $\mathrm{Diff}(\mathbb{S}^1)$ [Mil84]. This function cannot be reached by any one-parameter subgroup, but can be made as close as we want to the identity by dimensioning ϵ and n. However, such deformations are very unlikely in our applications since the norm of the k-th derivative $\|f_{n,\epsilon}\|_{H^k}$ is increasing when k is going to infinity. Regularity is indeed a critical issue in image registration. Thus, it may be a good idea to actually exclude this type of diffeomorphisms from the space under consideration

In practice, there is also an issue with the discretization of the flow of velocity fields: we inevitably have a spatial discretization of the velocity fields (and of the deformations) on a grid. For LDDMM methods, there is an addition temporal discretization of the time varying velocity fields by a fixed number of time steps. These discretizations intrinsically limit the frequency of the deformations, and prevent very high spatial frequency diffeomorphisms to be reached both by the SVF and by the discrete LDDMM frameworks. Chapter **??** of this book provides an insight on such numerical issues in LDDMM while Chapter **??** builds on this low-pass behavior to explicitly limits the deformation frequency. These limitations remains to be compared to the ones of the SVFs.

5.6.2. SVF-Based Setting: Properties and Algorithm

In SVF-based methods, the deformation $\phi = \exp(v)$ is parametrized by the Lie group exponential of a smooth SVF $v : \Omega \to \mathbb{R}^3$, defined by the ODE:

$$\frac{\partial \phi(x,t)}{\partial t} = v(\phi(x,t)), \tag{5.16}$$

with initial condition $\phi(x, 0) = x$. This ODE defines a one parameter subgroup: $\phi_t(x) = \phi(x, t)$ since $\phi_{s+t}(x) = \phi(x, s)\,\phi(x, t) = \phi(x, s + t)$. The Lie group exponential is obtained at the parameter value $t = 1$, i.e. $\exp(v) = \phi(x) = \phi(x, 1)$. The one-parameter subgroup structure is the key element of the SVF setting, as it provides an efficient way to tackle important computational problems in image registration, such as:

- the numerical integration of spatial vector fields (computing the exponential),
- the efficient inversion of spatial transformations,
- the stable computation of differential quantities and quantification of volume changes,
- the composition of spatial transformations.

We illustrate in this section the related schemes proposed in the literature.

5.6.2.1. Exponential of an SVF

The one-parameter subgroup property guarantees that the exponential operation can be efficiently implemented as the composition of transformations:

$$\exp(v) = \phi(x, 1) = \phi(x, \tfrac{1}{2})\,\phi(x, \tfrac{1}{2}) = \exp(v/2)\exp(v/2). \tag{5.17}$$

This property is at the core of the generalization of the "scaling and squaring" integration scheme from matrices to SVF [ACPA06]. As a result, the ODE (5.16) can be effectively computed as the iterative composition of successive exponentials (Algorithm 5.1).

Algorithm 5.1 Scaling and Squaring for the Lie group exponential

1. Scaling step: choose n so that $2^{-n}v$ is "small".
2. Compute a first approximation: $\phi_0(x) \leftarrow \exp(2^{-n}v)(x) \approx x + 2^{-n}v(x)$.
3. Squaring step: For $k = 1$ to n do $\phi_k \leftarrow \phi_{k-1}\,\phi_{k-1}$.

The initial integration step (Step 2 of Algorithm 5.1) is a sensitive step affecting the quality of the integration. This issue was investigated for example in [FLD+16], where different integration schemes and initial approximations strategy were benchmarked in terms of integration accuracy and computation time. Among the tested method, optimal performances were obtained with Runge-Kutta and Exponential Integrators methods [MVL03]. Moreover, computing the transformation $\phi_k(x_i) = \phi_{k-1}(\phi_{k-1}(x_i))$ at the image grid point x_i in the squaring step (Step 3 of Algorithm 5.1) involves a resampling since $\phi_{k-1}(x_i)$ is in general not at an exact grid point. The numerical accuracy of this step is critical. For instance, tri-linear interpolation can easily be non-diffeomorphic for large deformations.

5.6.2.2. Inversion of Spatial Transformations

Given a diffeomorphism ϕ, the computation of the inverse requires the estimation of a spatial transformation ϕ^{inv} such that $\phi(\phi^{inv}(x)) = x$, and $\phi^{inv}(\phi(x)) = x$. In general, this is performed using a least-squares optimization minimizing the error over the domain. In the SVF setting, such a computation can be performed through very simple algebraic manipulations on the related SVF parameters. Indeed, the one-parameter subgroup properties guarantees $\phi^{inv} = \phi(x, -1)$, since $\phi(x, 1)\phi(x, -1) = \phi(x, 0) = e$, and vice versa. Thus, the diffeomorphism $\phi = \exp(v)$ parametrized by the SVF v has inverse $\phi^{inv} = \exp(-v)$ parametrized by the SVF $-v$.

5.6.2.3. Computing differential quantities

The quantification of the amount of warping ϕ applied at each voxel by the dense deformation field is usually locally derived from the Jacobian matrix $d\phi = \nabla\phi^\top$ of the deformation in terms of determinant, log-determinant, trace, and the right Cauchy–Green strain tensor $d\phi^\top d\phi$. An index of volume change over an anatomical region Ω can be obtained by integrating the change of volume of all elements in the region of interest, or by computing the flux that is going inward or outward of this region.

Integration of the Jacobian Determinant in the Region of Interest

This is an average measure of *volume change*. The computation of the Jacobian matrix $d\phi$ is normally performed by spatial differentiation of the transformation by finite differences (Algorithm 5.2) .

Algorithm 5.2 Classical Computation of the Jacobian Determinant by Finite Differences

Given a discrete sampling ϕ of the transformation over the image grid space $\{x_i\}$:

1. Compute the Jacobian matrix F via finite differences along coordinates e_k:
   $d\phi_{jk}(x_i) = [\partial_{e_k}\phi(x_i)]_j = (\phi_j(x_i + \delta e_k) - \phi_j(x_i))/\delta$, where δ is the discretization step size.
2. Compute $\det(d\phi)$ with the preferred numerical method.

The differentiation by finite differences is however usually highly sensitive to the spatial noise, and completely depends on the size of the discrete space-sampling. This can create instabilities in case of large deformations leading to incorrect Jacobian determinant estimation. This limitation is elegantly overcome in the SVF framework using a variation of the scaling and squaring method for the Lie group exponential. In fact, the (log-) Jacobian can be reliably estimated by finite differences for the *scaled* velocity field $v/2^n$, and then recursively computed thanks to the additive property of the one-parameter subgroups, and by applying the chain rule (Algorithm 5.3).

Algorithm 5.3 Jacobian Matrix and Log-Jacobian Determinant with Scaling and Squaring

Given a deformation $\phi = \exp(v)$:

1. Scaling step: Choose n so that $2^{-n}v$ is "small".
2. Compute a first approximation (dv is computed using finite differences):
   $\phi_0 = \exp(2^{-n}v) \approx \text{Id} + 2^{-n}v \quad \text{and} \quad d\phi_0 \approx \text{Id} + 2^{-n}dv.$
   $LD_0 = \log\det(d\phi_0) = \text{Tr}(\log(d\phi_0)) \approx \text{Tr}(2^{-n}dv).$
3. Squaring step: For $k = 1$ to n do:
   $\phi_k = \phi_{k-1}\phi_{k-1} \quad \text{and} \quad d\phi_k = (d\phi_{k-1}\phi_{k-1})\, d\phi_{k-1}$
   $LD_k = \log\det(d\phi_k) = LD_{k-1} \circ \phi_{k-1} + LD_{k-1}.$

With this scheme the Jacobian determinant is therefore evaluated accordingly to the exponential path and is consistent with the definition of diffeomorphisms parametrized by the one-parameter subgroup. Moreover, the log-Jacobian determinant is defined in terms of the divergence of the velocity and, by definition, the value of the corresponding Jacobian determinant is always strictly positive. This property implies that the evaluation preserves the diffeomorphic formulation and is therefore robust to the discretization approximations. For instance, in case of large deformations, the sampling of the deformation field in the image grid space may introduce spurious folding effects (e.g. in presence of an unequal distribution of the vectors around a sink), thus leading to an incorrect negative Jacobian estimation with the direct estimation, while it would be still correctly defined with the present method. One should be careful that the estimation of the full Jacobian matrix $d\phi$ with this method is numerically less stable than the log-Jacobian as one need to resample and multiply many matrices that are very small deviations from the identity.

Flux of the Deformation Field Across the Boundary of the Region

We can also derive a volume change index by comparing the volume enclosed by the deformed surface relatively to the original one. This is related to the flux of the velocity field across the boundary of the region $\delta\Omega$. Nevertheless, the direct computation of the flux is usually hindered by its high sensitivity to the localization and orientation of the boundaries. This limitation led to the development of surrogate intensity-based measures of the flux [FF97, SDSJM01]. However, the properties of the SVF framework enables us to derive a stable and efficient numerical scheme. Following [LAFP13], formula (5.16) leads to the following relationship:

$$\int_{\Omega} \log(\det(d\phi(x,1)))d\Omega \quad = \quad \int_0^1 \mathrm{flux}_{\partial\Omega}(v \circ \phi(x,t))\, dt. \tag{5.18}$$

This formula shows that the spatial integration of the log-Jacobian determinant of the deformation over the region of interest is equal to the flux of the velocity field across the corresponding boundary, integrated along the path described by the exponential map. Formula (5.18) consistently computes the flow of the vector field during the evolution described by the SVF parametrization, and measures the flux of a vector field over a surface (right side of (5.18)) by scalar integration of the log-Jacobian determinant in the enclosed volume (left side of (5.18)). Moving from the surface to the volume integration simplifies and robustifies the measure of the flux by attenuating the segmentation errors (and relative erroneous boundary detection). This also allows to deal with uncertainties in the region of interest, for instance by integration on probabilistic masks. The difference between the Jacobian and the log-Jacobian analysis becomes clear: the former quantifies volume changes, while the latter quantifies the shift of the boundaries (given by the average regional log-Jacobian determinant).

5.6.2.4. Composing Transformations Parametrized by SVF

The Baker Campbell Hausdorff (BCH) formula (Section 5.3.1) was introduced in the SVF diffeomorphic registration in [BHO07] and provides an explicit way to compose diffeomorphisms parametrized by SVFs by operating in the Lie Algebra only. More specifically, if v, u are SVFs, then $\exp(v)\exp(u) = \exp(w)$ with

$$w = BCH(v,u) = v + u + \frac{1}{2}[v,u] + \frac{1}{12}[v,[v,u]] - \frac{1}{12}[u,[v,u]] + \dots$$

In this formula, the Lie bracket of vector fields $[v,u]$ is the derivative of u in the direction of the flow of v: $[v,u] = dv\, u - du\, v = \partial_u v - \partial_v u$. Here, the Lie algebra of diffeomorphisms is by convention the algebra of right-invariant vector fields instead of the traditional left-invariant ones used in finite dimensional Lie groups. This explains why this bracket is the opposite of the one of Section 5.3.1 (see comments in [VPPA08, BHO07]).

For a small u, the computation can be truncated at any order to obtain an approximation for the composition of diffeomorphisms. For this reason, the BCH is a key tool for the development of efficient gradient-based optimization algorithms. Keeping the description of our deformations within the Lie algebra significantly simplifies the optimization of the SVF parameters via gradient descent in the log-Demons algorithm.

5.6.3. SVF-Based Diffeomorphic Registration with the Log-Demons

Inspired by the idea of encoding diffeomorphisms with the flow of SVF [ACPA06], several SVF-based non-linear image registration algorithms were concurrently proposed [VPPA07, BHO07, Ash07, VPPA08, VPPA09, HBO09, MRD+11]. Among them, the (log)-Demons registration algorithm [VPPA07, VPPA08, VPPA09] found a considerable interest in the medical image registration community. Successful application to several clinical problems include [PDSP08, MPS+11, LFAP11, SPR11]. This setting is particularly appealing since it leads to a computationally effective and flexible registration scheme leveraging on the mathematics and numerics of SVF. We illustrate is this section the SVF properties presented in Section 5.6.2 within the log-Demons registration framework.

Image Similarity in the Log-Demons

Given a pair of images $I, J : \mathbb{R}^3 \mapsto \mathbb{R}$, we aim at estimating a SVF v parametrizing diffeomorphically the spatial correspondences that minimize a functional $Sim[I, J, v]$. For example, if the similarity is the log-likelihood of a Gaussian intensity error (the *sum of squared differences* criterion -SSD-), we may have $Sim[I, J, v] = \|I - J \circ \exp(-v)\|^2_{L_2}$, or $Sim[I, J, v] = \|I \circ \exp(v) - J\|^2_{L_2}$ depending on the choice of the reference image. The SVF formulation easily allows to symmetrize the similarity term in order to make it independant of the choice of the reference image. For example, [VPPA08] proposed unbiased correspondences by averaging the forward and backward correspondences $v = \frac{1}{2}(u + w)$ separately estimated from the SSD functional on both sides: $Sim_{forw}[I, J, u] = \|I \circ \exp(u) - J\|^2_{L_2}$ and $Sim_{back}[I, J, w] = \|I - J \circ \exp(-w)\|^2_{L_2}$. Although the symmetrization comes straightforwardly from the SVF parametrization of the deformations, the strategy requires twice the optimization of the correspondence terms, and can be computationally costly when extended to similarity terms more complex than the standard SSD.

To address this issue, [LAFP13] proposed a symmetric criterion optimizing at the half-way space, were both images are resampled simultaneously. This can be easily formulated with the SVF framework thanks to the inverse property by considering the resampled images $I \circ \exp(v/2)$ and $J \circ \exp(-v/2)$. For instance, the standard SSD can be symmetrized in:

$$SSD_{[\mathrm{sym}]}(I, J, v) = \|I \circ \exp(v/2) - J \circ \exp(-v/2)\|^2_{L_2}.$$

More complex similarity functionals, such as the local correlation coefficient (LCC), were easily extended to symmetric criteria using this formulation [LAFP13].

Regularization of SVF Parameters.
In order to prevent overfit, image registration usually considers a regularization term $Reg(v)$ aiming at promoting the spatial regularity of the solution. Several regularization functionals have been proposed in the literature to promote specific mechanistic constraints, such as diffusion properties $Reg(v) = \|dv\|^2_{L_2}$, incompressibility $Reg(v) = \|\mathrm{Tr}(dv)\|^2_{L_2}$ [MPS+11, MPM+12], or more complex terms involving the penalization of the (potentially infinite) high-order derivative terms of the SVF [CA04].

However, instead of adding the regularization to the similarity term as classically done in image registration, it was observed in [CBD+03] that introducing an auxiliary variable for the correspondences with a coupling term in the demons criterion was providing a more efficient optimization. In the log-demons framework, this amounts to parametrize the image correspondences by the flow of a SVF v_c, and the coupling term by

$$Aux(v_c, v) = \|v_c - v\|^2_{L_2} \approx \|BCH(v_c, -v)\|^2_{L_2} = \|\log(\exp(v_c)\exp(-v))\|^2_{L_2}.$$

The criterion optimized by the log-demons is then:

$$E(v, v_c, I, J) = \tfrac{1}{\sigma_i^2} Sim(I, J, v_c) + \tfrac{1}{\sigma_x^2} Aux(v_c, v) + \tfrac{1}{\sigma_T^2} Reg(v). \tag{5.19}$$

5.6.4. Optimizing the Log-Demons Algorithm

The interest of the auxiliary variable is to decouple a non-linear and non-convex optimization into two optimizations which are respectively local and quadratic. The classical criterion is obtain at the limit when the typical scale of the error σ_x^2 between the transformation and the correspondences tends to zero. The minimization of (5.19) is alternatively performed with respect to the SVF parameters v_c and v in two steps:

- *Matching.* The correspondence energy: $E_{\mathrm{corr}}(v, v_c, I, J) = \frac{1}{\sigma_i^2} Sim(I, J, v_c) + \frac{1}{\sigma_x^2} Aux(v_c, v)$, is minimized to find a (non-regularized) SVF v_c that best puts into correspondence the two images. The optimization of this non-convex energy is usually performed via gradient descent, Gauss-Newton or Levenberg Marquardt methods. Thanks to the quadratic formulation of the auxiliary term in v_c, the correspondence energy update can be efficiently computed with respect to standard similarity functionals $Sim(\cdot)$, such as the sum of squared differences (SSD), or the local correlation coefficient (LCC). In particular, the Taylor expansion of the similarity with respect to the variation δu of v_c leads to a closed form with a second order Newton's like gradient descent scheme:

$$\delta u = \left(\|\Lambda\|^2 + \frac{1}{Sim(I,J,v_c)} \frac{\sigma_i^2}{\sigma_x^2} \right)^{(-1)} \Lambda,$$

  where Λ is the gradient of $E_{\mathrm{corr}}(v, v_c, I, J)$ with respect the update δu [LAFP13].

- *Regularization.* The functional $E_{\mathrm{reg}}(v, v_c) = \frac{1}{\sigma_x^2} Aux(v_c, v) + \frac{1}{\sigma_T^2}\mathrm{Reg}(v)$ is optimized with respect to v. Following [MPS+11], formulating the term $Reg(\cdot)$ with infinite dimensional Isotropic Differential Quadratic Forms (IDQF, [CA04]) leads to a closed form for the regularization step: the optimal v is obtained using a Gaussian convolution $v = G_\sigma * v_c$, where σ is a parameter of the IDQF. Thus this regularization step can be solved explicitly and very efficiently.

5.7. Parallel Transport of SVF Deformations

Modeling the temporal evolution of the tissues of the body is an important goal of medical image analysis, for instance to understand the structural changes of organs affected by a pathology, or to study the physiological growth during the life span. This requires to analyze and compare the anatomical differences in time series of anatomical images of different subjects. The main difficulty is to compare across individuals the transformation parameters describing the anatomical changes over time (longitudinal deformations) within each subject.

Comparison of longitudinal deformations can be done in different ways, depending on the analyzed feature. For instance, the scalar Jacobian determinant of longitudinal deformations represents the associated local volume change, and can be compared by scalar resampling in a common reference frame via inter-subject registration. This simple transport of scalar quantities is the basis of the classical tensor based morphometry techniques [AF00, RLF+04]. However, transporting the Jacobian determinant is not sufficient to reconstruct a deformation in the template space.

If we consider vector-values characteristics of deformations instead of scalar quantities, the transport is not uniquely defined anymore. For instance, a simple method of transport consists in *reorienting* the longitudinal intra-subject displacement vector field by the Jacobian matrix of the subject-to-reference deformation. Another intuitive method uses the *transformation conjugation* (change of coordinate system) in order to compose the longitudinal intra-subject deformation with the subject-to-reference one [RCSO+04]. As pointed out in [BZO10], this method relies on the inverse consistency of the inter-subject deformations, which can raise numerical problems for large deformations. Among these normalization methods, the *parallel transport* of longitudinal deformations is arguably a more principled tool in the diffeomorphic registration setting thanks to its differential geometric background.

5.7.1. Continuous and Discrete Parallel Transport Methods

In computational anatomy, the parallel transport along geodesics of diffeomorphisms with a right invariant metric has been initially proposed in the LDDMM context by [You07]. This work builds upon the idea of approximating the parallel transport by Jacobi fields [Arn79]. An application of this framework can be found in [QYMC08] for the study of hippocampal shape changes in Alzheimer's disease. Although representing a rigorous implementation of the parallel transport, this framework is generally computationally intensive. Moreover, the formulation is quite specific to LDDMM geodesics, and it is not evident to extend it to a general computational scheme in affine spaces.

In the context of SVF-based registration, [LP13] provided explicit formulas for the parallel transport with respect to the standard Cartan-Schouten connections (left, right and symmetric) in the case of *finite dimensional Lie groups*. Although further investigations are needed to better understand the generalization to infinite dimensions, practical examples of parallel transport of longitudinal diffeomorphisms with respect to the Cartan-Schouten connections demonstrated that it was an effective approach to transport SVFs. However, experiments showed that the numerical implementation plays a central role in the stability and accuracy of the different transport methods. For instance, the left and symmetric Cartan transports appear to be less stable than the right one due to the need of computing high-order differentials.

The practical implementation of continuous transport methods requires a precise knowledge of connection underlying the space geometry. This is not always simple, especially in the image registration setting. Moreover, the parallel transport involves the computation of high-order derivatives which, in practice, may introduce numerical issues. In particular, differential operations are particularly sensitive to the discretization of energy functionals and operators on the image grid.

The complexity and limitations deriving from the direct computation of continuous parallel transport methods can be alleviated when considering *discrete approximations*. Inspired by the work of the theoretical physicist Alfred Schild's, the *Schild's Ladder* was proposed in [MTW73] as a scheme for performing the parallel transport through the construction of geodesic parallelograms. The interest of the Schild's ladder resides in the generality of its formulation, since it only requires to compute geodesics. This means in particular that this type of parallel transport remains consistent with the numerical scheme used to compute the geodesics. Indeed, although the geodesics on the manifold are not sufficient to recover all the information about the geometric properties of the space, such as the torsion of the connection, it was shown in [KMN00] that Schild's ladder approximates the parallel transport with respect to the symmetric part of the connection of the space at the first order.

5.7.2. Discrete ladders for the registration of image sequences

Let $\{I_i\}$ $(i = 1 \dots n)$ be a time series of images with the baseline I_0 as reference. The longitudinal deformation from I_0 to I_i can be computed with image registration, in our case encoded with SVFs. Our goal is to transport these registration parameters to a template image T_0 in order to produce follow-up image T_i that transforms the image sequence $I_0, \ldots, I_i$ to the corresponding sequence $T_0, \ldots T_i$ in the reference space. In order to relate geodesics between images to geodesics in the diffeomorphism group, we assume here that all images belong to the orbit $\mathbb{I}$ of the template (or of any other image of the sequence) and that our registration algorithm provides an exact matching. These are of course simplifying assumptions whose impact needs to be evaluated in practice. In particular, we know that the registration is never perfect as the regularization term always prevents from achieving a perfect image match.

Nevertheless, this assumption allows us to model the image space $\mathbb{I}$ as the quotient of the group of diffeomorphisms by the isotropy group of the template (this is the set of deformations that leaves the template image unchanged). This construction endows the image space $\mathbb{I}$ with an invariant affine structure. We assume here that the affine manifold manifold $(\mathbb{I}, \nabla)$ which is produced is a geodesically orbital space [AA07, NRS07] where all geodesics are homogeneous, i.e. orbits of one-parameter subgroups of deformations: $I(t) = I \circ \exp(tv)$ for $I \in \mathbb{I}$. This idealized setting is perfectly aligned with the SVF-based registration methodology[1].

5.7.2.1. Schild's Ladder

In the medical image registration domain, [LAP11, LP13] was the first to adapt Schild's ladder to SVF-based diffeomorphic image registration. Schild's ladder transports a vector along a curve through the construction of geodesic parallelograms. One step of the ladder is illustrated in Figure 5.1. The corresponding algorithm in our idealized image space model $\mathbb{I}$ is described in Algorithm 5.4.

Interestingly, the Schild's ladder implementation appeared to be more stable in practice than the closed-form expression of the symmetric Cartan-Schouten parallel transport on group geodesics of diffeomorphisms. The reason is probably the inconsistency of numerical schemes used for the computation of the geodesics and for the transformation Jacobian in the implementation of this exact parallel transport formula.

This scheme requires the computation of two logarithms/registrations (from I_1 to T_0, and from I_0 to $I_{1/2}$), and can thus be computationally expensive when transporting multiple images. Moreover, the transport of time series of $\{I_i\}$ images is not defined with respect to the same baseline-to-reference curve, since the mid-point $I_{1/2}$ depends on the follow-up image I_i (see Fig. 5.1). This issue is critical because it may introduce inconsistencies in practice. Theses problems may be tackled by modifying the Schild's ladder to lead to a novel scheme: the pole ladder.

[1]The study of this construction remains to be mathematically substantiated. It is notably complexified by the infinite dimension. However, we believe that this idealized setting is a good metaphor anyway because it provides an extraordinary simplifying framework to explain the parallel transport on images.

Algorithm 5.4 Schild's Ladder for the Transport of a Longitudinal Deformation.

Let I_0 and I_1 be a series of (two) images, and T_0 a reference frame.
1. Compute the geodesic $\gamma(s)$ in the space $\mathbb{I}$ connecting I_1 and T_0 and define the mid-point $I_{1/2} = \gamma(1/2)$.
2. Compute the geodesic $\rho(s)$ from I_0 to $I_{1/2}$ and shoot twice along this geodesic to define the transported follow-up image $T_1 = \rho(2)$.
3. The transported SVF $\log_{T_0}(T_1)$ is obtained by registering the images T_0 and T_1.

5.7.2.2. Pole Ladder

The pole ladder is a modified version of the Schild's Ladder based on the observation that if the curve along which we want to transport is itself a geodesic, then it can be used as one of the diagonals of the geodesic parallelogram. In this case, constructing the ladder for image time series requires the computation of a new diagonal of the parallelogram only, and is defined with respect to the same reference (Figure 5.1). The resulting ladder is therefore analogous to the Schild's one, with the difference of explicitly using as a diagonal the geodesic C which connects I_0 and T_0 (Algorithm 5.5).

Algorithm 5.5 Pole Ladder for the Transport of a Longitudinal Deformation.

Let I_0 and I_1 be a series of (two) images, and T_0 a reference frame.
1. Compute the geodesic $C(s)$ in the space $\mathbb{I}$ connecting I_0 and T_0 and define the mid-point $I_{1/2} = C(1/2)$.
2. Compute the geodesic $\gamma(t)$ from I_1 to $I_{\frac{1}{2}}$ and shoot twice along this geodesic to define the transported image $T_1 = \gamma(2)$.
3. The transported SVF is the inverse of velocity field registering T_0 to $p(1) = T_1'$.

5.7.2.3. Theoretical Accuracy: Pole Ladder is a Third Order Scheme

So far, Schild's and pole ladders methods were shown to be first order approximations of the Riemannian parallel transport. Building on a BCH-type formula on affine connection spaces, the behavior of one pole ladder step was recently established up to order 5 [Pen18]. In Lie groups, we have seen that the BCH formula provides an expansion of the composition of two group exponentials in the Lie algebra: $BCH(v, u) = \log(\exp(v)\ \exp(u))$. In general affine connection manifolds, a somewhat similar formula can be established based on the curvature instead of the Lie bracket: the double exponential $\exp_x(v, u) = \exp_y(\Pi_x^y u)$ corresponds to a first geodesic shooting from the

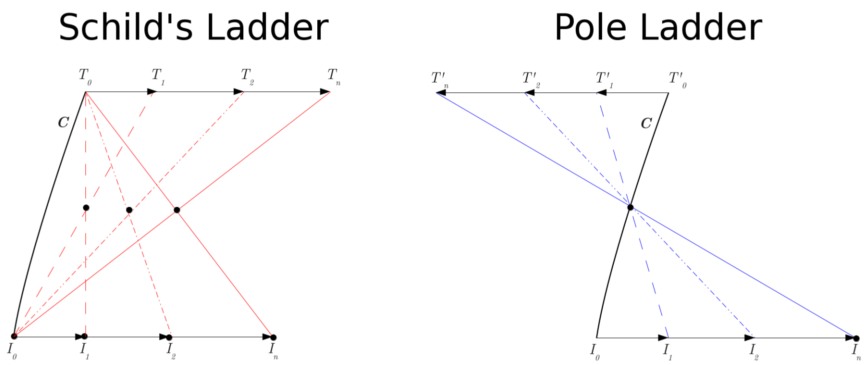

Figure 5.1 Geometrical schemes in the Schild's ladder and in the pole ladder. By using the curve C as diagonal, the pole ladder requires the computation of half times of the geodesics (blue) required by the Schild's ladder (red) (Figure adapted from [LP14]).

point x along the vector v, followed by a second geodesic shooting from $y = \exp_x(v)$ along the parallel transport $\Pi_x^y u$ of the vector u. [Gav06] has shown that the Taylor expansion of the log of this composition $h_x(v, u) = \log_x(\exp_x(v, u))$ is:

$$\begin{aligned} h_x(v,u) =& v + u + \frac{1}{6}R(u,v)v + \frac{1}{3}R(u,v)u + \frac{1}{12}\nabla_v R(u,v)v + \frac{1}{24}(\nabla_u R)(u,v)v \\ &+ \frac{5}{24}(\nabla_v R)(u,v)u + \frac{1}{12}(\nabla_u R)(u,v)u + O(\|u\|^5 + \|v\|^5). \end{aligned}$$

When applied to pole ladder reformulated using geodesic symmetry, we find that the error on one step of pole ladder to transport the vector u along the geodesic segment $[I_0, T_0] = [\exp_{I_{1/2}}(-v/2), \exp_{I_{1/2}}(v/2)]$ (all quantities being parallel translated at the mid-point $I_{1/2}$) is:

$$\Pi_{T_0}^{I_{1/2}}\mathrm{pole}(u) - \Pi_{I_0}^{I_{1/2}}u = \frac{1}{48}\left((\nabla_v R)(u,v)(5u - v) + (\nabla_u R)(u,v)(v - 2u)\right) + O(\|v + u\|^5).$$

It is remarkable that the scheme is of order three in general affine connection spaces with a symmetric connection, much higher than expected. Moreover, the fourth order error term vanishes in affine symmetric spaces since the curvature is covariantly constant in these spaces. In fact, one can actually prove that all error terms vanish in a convex normal neighborhood of an affine connection space: one step of pole ladder realizes a transvection, which is an exact parallel transport (provided that geodesics and mid-points are computed exactly of course) [Pen18]. The scheme is even globally exact in Riemannian symmetric manifolds. These properties make pole ladder a very attractive alternative for parallel transport in more general affine or Riemannian manifolds. In particular, pole ladder is exact for SVF: $\Pi^v(u) = \log(\exp(v/2)\exp(u)\exp(-v/2))$

(see Fig.5.2), as already noted in [LP14].

5.7.2.4. Effective Ladders on SVF-Deformations

Despite the straightforward formulation, algorithms (5.4) and (5.5) require multiple evaluations of geodesics in the space of diffeomorphisms, thus resulting in high computational cost. Moreover, since an exact matching is practically impossible, the implementation of the ladders through multiple image registrations may lead to important approximations of the parallel transport. For instance, the definition of $I_{1/2}$ using the forward deformation from I_0 or the backward one from T_0 may provide significantly different results. Finally, numerical approximations introduced by exponential and logarithm maps can introduce errors that can propagate during the iteration of the ladder. For all of these reasons, it is desirable to reformulate the above schemes using only transformations to obtain a computationally efficient and numerically stable framework.

Within the SVF framework, the transport can be very effectively approximated using the BCH formula for compositions [LP14]. The BCH approximation of the above exact parallel transport of a deformation $\exp(u)$ along a geodesic parametrized by the SVF v is obtained by:

$$\Pi^{v}_{BCH}(u) \quad \simeq \quad u + [v/2, u] + \frac{1}{2}[v/2, [v/2, u]] + HOT. \tag{5.20}$$

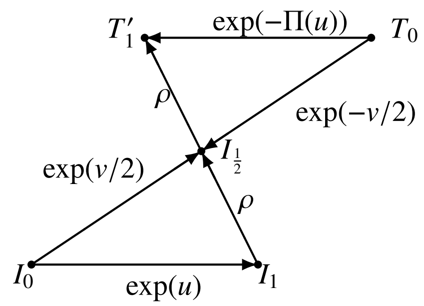

Figure 5.2 Ladder with the one parameter subgroups. The transport $\exp(\Pi(u))$ is the deformation $\exp(v/2)\exp(u)\exp(-v/2)$ (Figure adapted from [LP14]).

To gain even more efficiency and numerical stability, we can derive an iterative scheme for the computation of formula (5.20), inspired by the one-parameter subgroups properties of SVF. To provide a sufficiently small vector in the computation of the conjugate we observe that:

$$\exp(v)\exp(u)\exp(-v) = \exp\left(\frac{v}{n}\right)\ldots\exp\left(\frac{v}{n}\right)\exp(u)\exp\left(-\frac{v}{n}\right)\ldots\exp\left(-\frac{v}{n}\right).$$

The parallel transport can then be recursively computed by iterating the ladder over small geodesics parametrized by v/n:

1. Scaling step: find n such that v/n is small.

2. Iterate n times the ladder step: $u \leftarrow u + [\frac{v}{n}, u] + \frac{1}{2}[\frac{v}{n}, [\frac{v}{n}, u]]$.

We note that this method preserves the original "ladder" formulation, operated along the inter-subject geodesic $\exp(tv)$. In fact it iterates the construction of the ladder along the path $\exp(v)$ over small steps of size v/n.

5.7.3. Longitudinal Analysis of Brain Deformations in Alzheimer's Disease

We illustrate in this Section an application of pole ladder to the estimation of a group-wise model of the longitudinal changes in a group of patients affected by Alzheimer's disease (AD). In this disease, the brain atrophy which is measurable in time sequences of magnetic resonance images (MRI) was shown to strongly correlate with cognitive performance and neuropsychological scores, and characterizes the progression from pre-clinical to pathological stages [FFJJ+10]. For this reason, the development of reliable atlases of the pathological longitudinal evolution of the brain is of great importance for improving the understanding of the pathology. Tackling this problem requires the development of frameworks allowing the comparison across individuals of the atrophy trajectory measured through non-rigid registration.

A preliminary approach to the group-wise analysis of longitudinal morphological changes in AD consists in performing the longitudinal analysis after normalizing the anatomical images to a template space. A key issue here is the different nature of the changes occurring at the intra-subject level, reflecting the individual's atrophy over time, and the changes across different subjects, which are usually of larger magnitude and not related to specific biological process. To improve the quantification of the longitudinal dynamics, the intra-subject changes should be modeled at the individual level, and only subsequently transported in the common reference for statistical analysis. For this reason, the parallel transport of longitudinal deformations is an ideal tool for the comparison of longitudinal trajectories, allowing statistical analysis of the longitudinal brain changes in a common reference frame. We summarize below the findings of [LP14].

Data Analysis and Results [LP14]

Images corresponding to the baseline I_0 and the one-year follow-up I_1 scans were selected for 135 subjects affected by Alzheimer's disease. For each subject i, the pairs of scans were rigidly aligned. The baseline was linearly registered to an unbiased reference template and the parameters of the linear transformation were applied to I_1^i. Finally, for each subject, the longitudinal changes were measured by non-linear registration using the LCC-Demons algorithm [LAFP13].

The resulting deformation fields $\phi_i = \exp(v_i)$ were transported with the pole ladder (BCH scheme) in the template reference along the non-linear subject-to-template deformation. The group-wise longitudinal progression was modeled as the mean of the transported SVFs v_i. The areas of significant longitudinal changes were investigated by one-sample t-test on the group of log-Jacobian scalar maps corresponding to the transported deformations, in order to detect the areas of measured expansion/contraction significantly different from zero.

For the sake of comparison, the one sample t-statistic was tested on the individual's longitudinal log-Jacobian scalar maps warped into the template space along the subject-to-template deformation. This is the classical transport used in tensor-based morphometry studies [AF00].

Figure 5.3 illustrates the mean SVF of the transported one-year longitudinal trajectories. The field flowing outward of the ventricles indicates a pronounced enlargement. Moreover, we notice an expansion in the temporal horns of the ventricles as well as a consistent contracting flow in the temporal areas. The same effect can be statistically quantified by evaluating the areas where the log-Jacobian maps are significantly different from zero. The areas of significant expansion are located around the ventricles and spread in the CSF, while a significant contraction is appreciable in the temporal lobes, hippocampi, parahippocampal gyrus and in the posterior cingulate. The statistical result is in agreement with the one provided by the simple scalar interpolation of the individual's longitudinal log-Jacobian maps. In fact we do not experience any substantial loss of localization power by transporting SVFs instead of scalar log-Jacobian maps. However by parallel transporting we preserve also the multidimensional information of the SVFs, which potentially leads to more powerful voxel-by-voxel comparisons than the ones obtained with univariate tests on scalars. For instance, we were able to show statistically significant different brain shape evolutions depending on the level of $A\beta_{1-42}$ protein in the CSF which could be pre-symptomatic of Alzheimer's disease [LFAP11]. More generally, a normal longitudinal deformation model allows to disentangle normal aging component from the pathological atrophy even with one time-point only per patient (cross-sectional design) [LPFA14].

The SVF describing the trajectory transported in a common template can also be decomposed into local volume changes and a divergence free reoriention pattern using Helmholtz' decomposition [LAP15]. This allows to consistently define anatomical regions of longitudinal brain atrophy in multiple patients, leading to improved measurements of the quantification of the longitudinal hippocampal and ventricular atrophy in AD. The method provided best performing results during the MIRIAD atrophy chal-

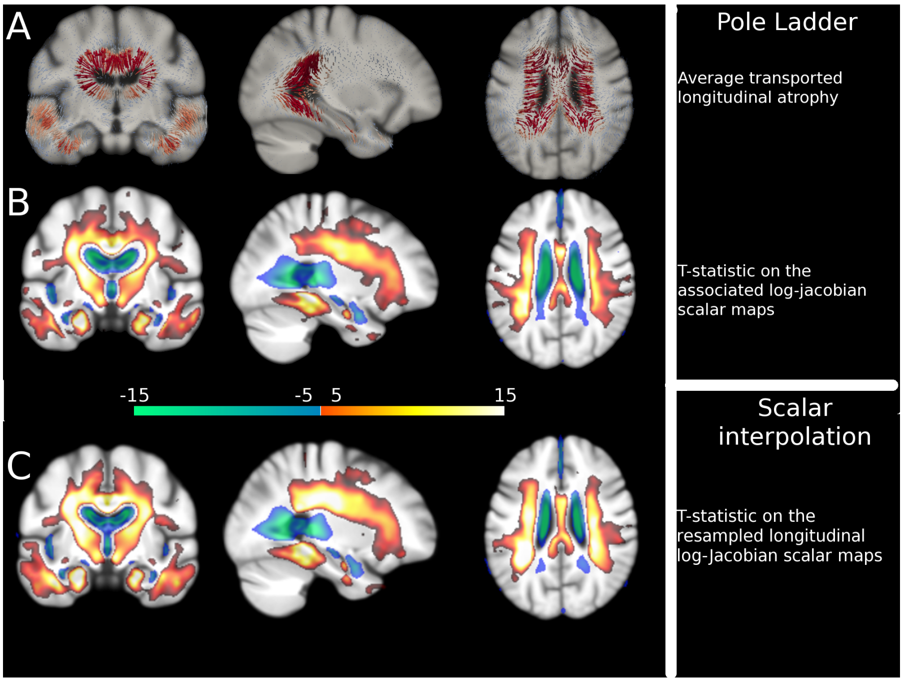

Figure 5.3 One year structural changes for 135 Alzheimer's patients. A) Mean of the longitudinal SVFs transported in the template space with the pole ladder. We notice the lateral expansion of the ventricles and the contraction in the temporal areas. B) T-statistic for the corresponding log-Jacobian values significantly different from 0 ($p < 0.001$ FDR corrected). C) T-statistic for longitudinal log-Jacobian scalar maps resampled from the subject to the template space. Blue color: significant expansion, Red color: significant contraction (Figure reproduced from [LP14]).

lenge for the regional atrophy quantification in the brain, with a favorable comparison with respect to state-of-art approaches [CFI$^+$15].

5.8. Historical Notes and Additional References

A large part of the body of sections 5.2 and 5.3 is based on the standard books on Riemannian manifolds [Kli82, dC92, GHL93]. Most parts related to the Cartan connection are taken from the book of Helagson [Hel78] which is the absolute reference for Lie groups. However, notations and a number of coordinate-free formulations are taken from more modern books on differential geometry and Lie groups. Among them [Pos01] is certainly one of the clearest exposition, especially for the affine setting (Chapters 1 to 6). On the link between left or right invariant geodesics on infinite dimensional Lie groups and Mechanics, the presentation of Kolev [Kol07] is really enlightening.

The barycentric definition of bi-invariant means on Lie groups based on one-parameter subgroups was developed during the PhD of Vincent Arsigny [Ars06] and in the research report [APA06]. In this preliminary work, the 'group geodesics' were simply defined as left translations of one-parameters subgroups without further justification. [PA12] extended this work by reformulating and rigorously justifying 'group geodesics' as the geodesics of the canonical Cartan-Schouten connections. This allows better distinguishing the properties that are related to the connection itself (bi-invariance) from the ones that are related to the definition of the mean as an exponential barycenter in an affine connection space.

Bibliography

[AA07] Dmitri Alekseevsky and Andreas Arvanitoyeorgos. Riemannian flag manifolds with homogeneous geodesics. *Transactions of the American Mathematical Society*, 359(8):3769–3789, 2007.

[ACAP09] Vincent Arsigny, Olivier Commowick, Nicholas Ayache, and Xavier Pennec. A fast and log-Euclidean polyaffine framework for locally linear registration. *J. of Math. Imaging and Vision (JMIV)*, 33(2):222–238, 2009.

[ACPA06] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-Euclidean framework for statistics on diffeomorphisms. In *Proc. of the 9th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'06), Part I*, number 4190 in LNCS, pages 924–931, 2-4 October 2006. PMID: 17354979.

[AF00] John Ashburner and Karl J Friston. Voxel-based morphometry - the methods. *Neuroimage*, 11(6):805–821, 2000.

[AL05] Marc Araudon and Xue-Mei Li. Barycenters of measures transported by stochastic flows. *The annals of probability*, 33(4):1509–1543, 2005.

[APA06] Vincent Arsigny, Xavier Pennec, and Nicholas Ayache. Bi-invariant means in lie groups. application to left-invariant polyaffine transformations. Research report rr-5885, INRIA Sophia-Antipolis, April 2006.

[Arn79] Vladimir Igorevich Arnol'd. *Mathematical methods of classical mechanics*. Springer Science & Business Media, 1979.

[Arn94] Marc Arnaudon. Espérances conditionnelles et C-martingales dans les variétés. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXVIII*, volume 1583 of *Lect. Notes in Math.*, pages 300–311. Springer-Verlag, 1994.

[Arn95] Marc Arnaudon. Barycentres convexes et approximations des martingales continues dans les variétés. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXIX*, volume 1613 of *Lect. Notes in*

Math., pages 70–85. Springer-Verlag, 1995.

[Ars06] Vincent Arsigny. *Processing Data in Lie Groups: An Algebraic Approach. Application to Non-Linear Registration and Diffusion Tensor MRI*. Thèse de sciences (phd thesis), École polytechnique, November 2006.

[Ash07] John Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95 – 113, 2007.

[BBM14] Martin Bauer, Martins Bruveris, and Peter W. Michor. Uniqueness of the Fisher-Rao metric on the space of smooth densities. *ArXiv e-prints*, November 2014.

[BHO07] Matias Bossa, Monica Hernandez, and Salvador Olmos. Contributions to 3D diffeomorphic atlas estimation: Application to brain images. In Nicholas Ayache, Sébastien Ourselin, and Anthony Maeder, editors, *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2007)*, volume 4792 of *LNCS*, pages 667–674. Springer-Verlag, 2007.

[BK81] Peter Buser and Hermann Karcher. *Gromov's almost flat manifolds*. Number 81 in Astérisque. Société mathématique de France, 1981.

[BMMTY05] Mirza Faisal Beg, Michael I. M.I. Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. Journal of Computer Vision*, 61(2):139–157, 2005.

[BZO10] Mathtias Bossa, Ernesto Zacur, and Salvador Olmos. On changing coordinate systems for longitudinal tensor-based morphometry. In *Proc. of Spatio Temporal Image Analysis Workshop (STIA 2010)*, page 44, 2010.

[CA04] Pascal Cachier and Nicholas Ayache. Isotropic energies, filters and splines for vector field regularization. *Journal of Mathematical Imaging and Vision*, 20(3):251–265, 2004.

[CBD+03] Pascal Cachier, Eric Bardinet, Didier Dormont, Xavier Pennec, and Nicholas Ayache. Iconic feature based nonrigid registration: The pasha algorithm. *Computer Vision and Image Understanding*, 89(2-3):272–298, Feb.-march 2003. Special Issue on Nonrigid Registration.

[CFI+15] David M. Cash, Chris Frost, Leonardo O. Iheme, Devrim Ünay, Melek Kandemir, Jurgen Fripp, Olivier Salvado, Pierrick Bourgeat, Martin Reuter, Bruce Fischl, Marco Lorenzi, Giovanni B. Frisoni, Xavier Pennec, Ronald K. Pierson, Jeffrey L. Gunter, Matthew L. Senjem, Clifford R. Jack, Nicolas Guizard, Vladimir S. Fonov, D. Louis Collins, Marc Modat, M. Jorge Cardoso, Kelvin K. Leung, Hongzhi Wang, Sandhitsu R. Das, Paul A. Yushkevich, Ian B. Malone, Nick C. Fox, Jonathan M. Schott, and Sebastien Ourselin. Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge. *NeuroImage*, 123:149–164, December 2015.

[CK99] José Manuel Corcuera and Wilfrid S. Kendall. Riemannian barycentres and geodesic convexity. *Math. Prob. Camb. Phil. Soc*, 127:253–269, 1999.

[CS26] Elie Cartan and Jan Arnoldus Schouten. On the geometry of the group-manifold of simple and semi-simple groups. *Proc. Akad. Wekensch, Amsterdam*, 29:803–815, 1926.

[dC92] Manfredo do Carmo. *Riemannian Geometry*. Mathematics. Birkhäuser, Boston, Basel, Berlin, 1992.

[EM91] Michel Emery and Gabriel Mokobodzki. Sur le barycentre d'une probabilité dans une variété. In M. Yor J. Azema, P.A. Meyer, editor, *Séminaire de probabilités XXV*, volume 1485 of *Lect. Notes in Math.*, pages 220–233. Springer-Verlag, 1991.

[FF97] Peter A Freeborough and Nick C Fox. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE transactions on medical imaging*, 16(5):623–629, 1997.

[FFJJ+10] Giovanni B Frisoni, Nick C Fox, Clifford R Jack Jr, Philip Scheltens, and Paul M Thompson. The clinical use of structural MRI in Alzheimer's disease. *Nature Reviews Neurology*, 6(2):67, 2010.

[FLD+16] Sebastiano Ferraris, Marco Lorenzi, Pankaj Daga, Marc Modat, and Tom Vercauteren. Accurate small deformation exponential approximant to integrate large velocity fields: Application to image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 17–24, 2016.

[Gal08] Jean Gallier. Logarithms and Square Roots of Real Matrices. *arXiv:0805.0245 [math]*, May 2008. arXiv: 0805.0245.

[Gav06] Aleksei Vladimirovich Gavrilov. Algebraic properties of covariant derivative and composition of exponential maps. *Matematicheskie Trudy*, 9(1):3–20, 2006.

[GHL93] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine. *Riemannian Geometry*. Springer-Verlag, 2nd edition edition, 1993.

[God82] Roger Godement. *Introduction à la Théorie des Groupes de Lie, Tomes I et II.* Publications Mathématiques de l'Université Paris VII, 1982.

[Gro04] David Groisser. Newton's method, zeroes of vector fields, and the Riemannian center of mass. *Adv. in Applied Math*, 33:95–135, 2004.

[HBO09] Monica Hernandez, Matias Bossa, and Salvador Olmos. Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision*, 85:291–306, 2009.

[HCHKL01] Sheung Hun Cheng, Nicholas J. Higham, Charles S. Kenney, and Alan J. Laub. Approximating the logarithm of a matrix to specified accuracy. *SIAM J. Matrix Anal. Appl.*, 22(4):1112–1125, 2001.

[Hel78] Sigurdur Helgason. *Differential Geometry, Lie groups, and Symmetric Spaces*. Academic Press, 1978.

[Hig05] Nicholas J. Higham. The scaling and squaring method for the matrix exponential revisited. *SIAM Journal on Matrix Analysis and Applications*, 26(4):1179–1193, January 2005.

[HLW02] Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration : structure preserving algorithm for ordinary differential equations*, volume 31 of *Springer series in computational mathematics*. Springer, 2002.

[IMKNZ00] Arieh Iserles, Hans Z. Munthe-Kaas, Syvert P. Norsett, and Antonella Zanna. Lie-group methods. *Acta numerica*, 9:215–365, 2000.

[Ken91] Wilfrid S. Kendall. Convexity and the hemisphere. *Journal of the London Mathematical Society*, 43(2):567–576, 1991.

[Ken92] Wilfrid S. Kendall. The propeller: a counterexample to a conjectured criterion for the existence of certain harmonic functions. *Journal of the London Mathematical Society*, 46:364–374, 1992.

[KL89] Charles S. Kenney and Alan J. Laub. Condition estimates for matrix functions. *SIAM J. Matrix Anal. Appl.*, 10:191–209, 1989.

[Kli82] Wilhelm Klingenberg. *Riemannian Geometry*. Walter de Gruyter, Berlin, New York, 1982.

[KMN00] Arkady Kheyfets, Warner A Miller, and Gregory A Newton. Schild's ladder parallel transport procedure for an arbitrary connection. *International Journal of Theoretical Physics*, 39(12):2891–2898, 2000.

[Kol07] Boris Kolev. Groupes de lie et mécanique. http://www.cmi.univ-mrs.fr/ kolev/, 2007. Notes of a Master course in 2006-2007 at Université de Provence.

[KW09] Boris A. Khesin and Robert Wendt. *The Geometry of Infinite Dimensional Lie groups*, volume 51 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics*. Springer Verlag, 2009.

[LAFP13] Marco Lorenzi, Nicholas Ayache, Giovanni B. Frisoni, and Xavier Pennec. LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*, 81(1):470–483, 2013.

[Lan04] Serge Lang. *Algebra*. Graduate Texts in Mathematics. Springer, 3rd rev. ed. 2002. corr. 4th printing edition, 2004.

[LAP11] Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. Schild's ladder for the parallel transport of deformations in time series of images. In G. Szekely and H. Hahn, editors, *IPMI - 22nd International Conference on Information Processing in Medical Images- 2011*, volume 6801, pages 463–474, Kloster Irsee, Germany, July 2011. Springer. Honorable Mention (runner-up) for the Erbsmann Award.

[LAP15] Marco Lorenzi, Nicholas Ayache, and Xavier Pennec. Regional flux analysis for discovering and quantifying anatomical changes: An application to the brain morphometry in Alzheimer's disease. *NeuroImage*, 115:224–234, July 2015.

[Laq92] H. Turner Laquer. Invariant affine connections on Lie groups. *Transactions of the American Mathematical Society*, 331(2):pp. 541–551, 1992.

[LFAP11] Marco Lorenzi, Giovanni B. Frisoni, Nicholas Ayache, and Xavier Pennec. Mapping the effects of $A\beta_{1-42}$ levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields. In G. Fichtinger, A. Martel, and T. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011*, volume 6893 of *LNCS*, pages 663–670. Springer, Heidelberg, Sep 2011.

[LP13] Marco Lorenzi and Xavier Pennec. Geodesics, parallel transport & one-parameter subgroups for diffeomorphic image registration. *International Journal of Computer Vision*, 105(2):111–127, November 2013.

[LP14] Marco Lorenzi and Xavier Pennec. Efficient parallel transport of deformations in time series of images: from Schild's to pole ladder. *Journal of Mathematical Imaging and Vision*, 50(1-2):5–17, 2014.

[LPFA14] Marco Lorenzi, Xavier Pennec, Giovanni B. Frisoni, and Nicholas Ayache. Disentangling normal aging from Alzheimer's disease in structural MR images. *Neurobiology of Aging*, September 2014.

[Med82] Alberto Medina. Groupes de Lie munis de pseudo-métriques de Riemann bi-invariantes. *Séminaire de géométrie différentielle 1981-1982*, 1982.

[Mil84] John Milnor. Remarks on infinite-dimensional Lie groups. In *Relativity, Groups and Topology*, pages 1009–1057. Les Houches, 1984.

[Moa02] Maher Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, January 2002.

[MP15] Nina Miolane and Xavier Pennec. Computing bi-invariant pseudo-metrics on Lie groups for consistent statistics. *Entropy*, 17(4):1850–1881, April 2015.

[MPM+12] Kristin McLeod, Adityo Prakosa, Tommaso Mansi, Maxime Sermesant, and Xavier Pennec. An incompressible log-domain demons algorithm for tracking heart tissue. In Oscar Camara, Ender Konukoglu, Mihaela Pop, Kawal Rhode, Maxime Sermesant, and Alistair Young, editors, *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*, volume 7085, pages 55–67. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[MPS+11] Tommaso Mansi, Xavier Pennec, Maxime Sermesant, Hervé Delingette, and Nicholas Ayache. iLogDemons: A demons-based registration algorithm for tracking incompressible elastic biological tissues. *International Journal of Computer Vision*, 92(1):92–111, 2011.

[MR85] Alberto Medina and Philippe Revoy. Algèbres de Lie et produit scalaire invariant. *Annales scientifiques de l'Ecole normale supérieure*, 18(3):553–561, 1985.

[MRD+11] Marc Modat, Gerard R. Ridgway, Pankaj Daga, Manuel Jorge Cardoso, David J. Hawkes, John Ashburner, and Sébastien Ourselin. Log-euclidean free-form deformation. In *Proc. of SPIE Medical Imaging 2011*. SPIE, 2011.

[MTW73] Charles W Misner, Kip S Thorne, and John Archibald Wheeler. *Gravitation*. W.H. Freeman and Company, 1973.

[MVL03] Cleve Moler and Charles Van Loan. Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM review*, 45(1):3–49, 2003.

[NRS07] Yu. G. Nikonorov, Evgenii Dmirtievich Rodionov, and Viktor Vladimirovich Slavskii. Geometry of homogeneous Riemannian manifolds. *Journal of Mathematical Sciences*, 146(6):6313–6390, November 2007.

[PA12] Xavier Pennec and Vincent Arsigny. Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups. In Frederic Barbaresco, Amit Mishra, and Frank Nielsen, editors, *Matrix Information Geometry*, pages 123–168. Springer, May 2012.

[PDSP08] Jean-Marc Peyrat, Hervé Delingette, Maxime Sermesant, and Xavier Pennec. Registration of 4D time-series of cardiac images with multichannel diffeomorphic demons. In D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, volume 5242 of *LNCS*, pages 972–979. Springer, Heidelberg, Sep 2008.

[Pen96] Xavier Pennec. *L'incertitude dans les problèmes de reconnaissance et de recalage – Applications en imagerie médicale et biologie moléculaire*. Phd thesis, Ecole Polytechnique, December 1996.

[Pen98] Xavier Pennec. Computing the mean of geometric features - application to the mean rotation. Research Report RR-3371, INRIA, March 1998.

[Pen06a] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006. A preliminary appeared as INRIA RR-5093, January 2004.

[Pen06b] Xavier Pennec. *Statistical Computing on Manifolds for Computational Anatomy*. Habilitation à diriger des recherches, Université Nice Sophia Antipolis, December 2006.

[Pen18] Xavier Pennec. Parallel transport with pole ladder: a third order scheme in affine connection spaces which is exact in affine symmetric spaces. *arXiv:1805.11436 [cs, math]*, May 2018. arXiv: 1805.11436.

[Pos01] Mikhail Mikhailovich Postnikov. *Geometry VI: Riemannian Geometry*. Encyclopedia of mathematical science. Springer, 2001.

[QYMC08] Anqi Qiu, Laurent Younes, Michael I Miller, and John G Csernansky. Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the alzheimer's type. *NeuroImage*, 40(1):68–76, 2008.

[RCSO+04] Anil Rao, Raghavendra Chandrashekara, Gerardo I Sanchez-Ortiz, Raad Mohiaddin, Paul Aljabar, Joseph V Hajnal, Basant K Puri, and Daniel Rueckert. Spatial transformation of motion and deformation fields using nonrigid registration. *IEEE Transactions on Medical Imaging*, 23(9):1065–1076, 2004.

[RLF+04] William R Riddle, Rui Li, J Michael Fitzpatrick, Susan C DonLevy, Benoit M Dawant, and Ronald R Price. Characterizing changes in mr images with color-coded jacobians. *Magnetic resonance imaging*, 22(6):769–777, 2004.

[SDSJM01] Stephen M Smith, Nicola De Stefano, Mark Jenkinson, and Paul M Matthews. Normalized accurate measurement of longitudinal brain change. *Journal of computer assisted tomography*, 25(3):466–475, 2001.

[SPR11] Christof Seiler, Xavier Pennec, and Mauricio Reyes. Geometry-aware multiscale image registration via OBBTree-based polyaffine log-demons. In G. Fichtinger, A. Martel, and T. Peters, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2011*, volume 6893 of *LNCS*, pages 631–638. Springer, Heidelberg, Sep 2011.

[Ste64] Shlomo Sternberg. *Lectures on Differential Geometry*. Prentice Hall Mathematics Series. Prentice Hall Inc., 1964.

[Tro98] Alain Trouvé. Diffeomorphisms groups and pattern matching in image analysis. *International Journal of Computer Vision*, 28(3):213–221, 1998.

[VPPA07] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas

Ayache. Non-parametric diffeomorphic image registration with the Demons algorithm. In Nicholas Ayache, Sébastien Ourselin, and Anthony Maeder, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007*, pages 319–326, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[VPPA08] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Symmetric Log-domain diffeomorphic registration: A Demons-based approach. In D. Metaxas, L. Axel, G. Fichtinger, and G. Szekely, editors, *Proc. of Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, volume 5241, pages 754–761. Springer, Heidelberg, Sep 2008.

[VPPA09] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72, March 2009.

[W03̈] Michael Wüstner. A connected Lie group equals the square of the exponential image. *Journal of Lie Theory*, 13:307–309, 2003.

[You07] Laurent Younes. Jacobi fields in groups of diffeomorphisms and applications. *Quarterly of applied mathematics*, pages 113–134, 2007.

[ZKC99] Milos Zefran, Vijai Kumar, and Christopher Croke. Metrics and connections for rigid-body kinematics. *Int. Journal of Robotics Research*, 18(2):243–258, February 1999.